**University of Alberta**

# Molecular systematics of the Arctoidea (Carnivora, Mammalia) and the use and utility of multiple gene datasets

by

Tara Lynn Fulton

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Systematics and Evolution

Department of Biological Sciences

Edmonton, Alberta

Fall 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*
*ISBN: 978-0-494-46318-5*
*Our file Notre référence*
*ISBN: 978-0-494-46318-5*

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

**Abstract**

Carnivores are closely tied to humans ecologically, economically, and emotionally. Despite these ties, much about carnivore evolution remains unknown. As technology improves, DNA sequence information from across the genome can be utilized to estimate carnivore evolutionary history (phylogeny). But as data increases, so, too, does the difficulty in modeling the various evolutionary processes of different genes. Using up to ~21,000 base pairs of new mitochondrial and nuclear DNA sequence per species, I addressed systematic questions for the arctoid carnivores while assessing the analysis techniques available for heterogeneous datasets. These include the importance of identifying discordance between genes and between analyses, properly partitioning a dataset into similarly evolving units, addressing evolutionary model violations, and investigating the strengths and weaknesses of current tree estimation methods.

Arctoid carnivores are comprised of three lineages: bears, pinnipeds, and musteloids (skunks, weasels, raccoons, red panda). Nuclear DNA analyses illustrated that the bears were the first of these three lineages to diverge. New musteloid familial relationships were proposed, where the skunk family diverged first, then the red panda, then the raccoon and weasel families. Using new nuclear gene sequences illustrated that the time between divergences was too short to accumulate many informative mutations, compounded by possible deep coalescent events. A new subfamily structure was proposed for the taxonomically problematic weasel family. Morphological convergence between ecologically similar procyonid genera (raccoons and relatives) was revealed through estimation of the first genus-level molecular phylogeny. Higher-level classification for the true seals was confirmed. Using multiple individuals per seal species

improved resolution of recent, rapid radiations, as one genus was recovered as monophyletic for the first time with molecular data. Accounting for base composition bias across taxa reconciled the nuclear and mitochondrial topologies and provided a new phylogenetic hypothesis for the Antarctic seals. A relaxed-clock molecular divergence time analysis using nuclear and mitochondrial data with multiple fossil calibrations allowed a revised biogeographic hypothesis for the true seals. Intrafamilial divergences were more recent than previously thought, indicating that northern seals may have had a long association with ice and that all southern seal tribes may have had a western Atlantic origin.

**Acknowledgements**

I would like to thank my supervisor, Dr. Curtis Strobeck, for allowing me the freedom and providing me the financial support to create the thesis I envisioned.

I also thank the other members of my supervisory committee, Dr. David Hik and Dr. David Coltman, and my candidacy and defense examiner, Dr. Stephen Moore for their time and contributions to bettering my thesis and learning. Thank you to Dr. Isabelle Delisle for bringing me into the field of carnivore phylogenetics, joining the defense examining committee on short notice, and much help in getting to this point – merci beaucoup. Special thanks to Dr. Felix Sperling for acting as an examiner for my candidacy, a graduate course instructor, and a great source of support. Finally, I thank Dr. Bill Murphy for his time and invaluable comments and contributions.

Invaluable critique on the content and writing of this thesis was provided by many friends and labmates: Greg Wilson, Lindsey Carmichael, Amanda Roe, Karen Cranston, Klaus Koepfli, Corey Davis, and Tara Klassen. Thank you! I believe that I am a better scientist because of knowing all of you. Karen, I couldn't have written Chapter 3 without you. Thanks for the technical genius and unintentional inspiration! Lindsey, Greg, and Amanda: thank you all for your much-needed support and guidance. And to Troy Locke who is always ready with a hug and a coffee – you are a great friend and comfort.

I would like to thank all of the distinguished staff that taught the UC Davis Bodega Bay phylogenetics workshop and the Woods Hole Workshop on Molecular Evolution. I gained a great amount of knowledge, perspective, and confidence from both experiences and I greatly appreciate these opportunities.

I could not have performed these studies without all of the people and institutions who provided me with necessary DNA samples. Though too numerous to name here, I thank you all and individual recognition is included in each published manuscript. Years of sequencing were made considerably easier through use of the departmental Molecular Biology Facility and the advice and support provided by the MBF staff.

I am grateful to the National Science and Engineering Research Council (NSERC), who made this research possible through both research and personal funding. I am also grateful to the Department of Biological Sciences and the University of Alberta for teaching assistantships and teaching and academic awards.

My entire graduate program was made infinitely easier by the unconditional love and unwavering support of my family and in-laws. You were all ready to listen at any time and I truly appreciate all of your efforts at reading jargon-filled manuscripts! I cannot thank my husband, Scott Fulton, enough for his patience, love, and support. You held me up, pulled me back down to reality, pushed me forward, and always stood beside me. Thank you all.

**Table of Contents**


**1. General Introduction** ..... **1**

The Carnivora ..... 1

Carnivore systematics ..... 2

*Carnivoramorpha* ..... 2

*Phylogeny and classification of extant Carnivora* ..... 2

Multiple gene analysis ..... 6

Thesis objectives ..... 7

References ..... 10

**2. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets** ..... **16**

Introduction ..... 16

Materials and methods ..... 19

*Samples, amplification, and sequencing strategy* ..... 19

*Phylogenetic analysis* ..... 20

Results ..... 22

*General sequencing and tree estimation results* ..... 22

*Relationships between Pinnipedia, Ursidae, and Musteloidea* ..... 23

*Species relationships within Ursidae* ..... 23

*Relationships within Pinnipedia* ..... 24

*Relationships within Musteloidea* ..... 25

*Effect of missing data: all STS vs. total data* ..... 27

*Supertree vs. supermatrix* ..... 28

Discussion ..... 29

*Higher-level systematics of the Arctoidea* ..... 29

*Relationships within Ursidae* ..... 30

*Relationships within Pinnipedia* ..... 30

*Musteloid family-level systematics: phylogenetic affinity of the red panda* ..... 31

*Species relationships within Mephitidae and Procyonidae* ..... 32

*Subfamily and species relationships within Mustelidae* ..... 33

*Effect of missing data: all STS vs. total data* ..... 35

*Supertree vs. supermatrix* ..... 36

References ..... 50

**3. Locus congruence and conflict and the effects of taxon sampling on elucidating the phylogenetic position of the red panda** ..... **54**

Introduction ..... 54

Materials and methods ..... 55

*Samples and sequencing* ..... 55
*Sequence alignment, marker congruence, and model selection* ..... 56
*Base composition bias and saturation* ..... 56
*Relative rate tests* ..... 57
*Tree search analyses* ..... 57
*Effects of Mephitidae taxon sampling and outgroup selection* ..... 58
*Hypothesis testing* ..... 58
Results ..... 59
*Sequencing results* ..... 59
*Locus congruence* ..... 60
*Base composition bias and saturation plots* ..... 60
*Rate variation across lineages* ..... 61
*Mitochondrial topologies* ..... 62
*Individual nuclear gene results and tests of alternate topologies* ..... 63
*Combined nuclear datasets: parsimony vs. minimum evolution vs. likelihood* ..... 64
*Combined nuclear and mtDNA* ..... 64
*Effects of Mephitidae taxon sampling and outgroup selection* ..... 64
*Outgroup relationships* ..... 65
Discussion ..... 66
*Unraveling a rapid radiation* ..... 66
*Taxon sampling and outgroup selection* ..... 67
*Identification of data biases* ..... 68
*Utility of mtDNA in resolving basal musteloid relationships* ..... 69
*Individual nuclear genes: support and conflict* ..... 70
*Mephitidae-basal: Is consistently strong support enough?* ..... 72
References ..... 90

**4. Novel multi-gene phylogeny of the raccoon family (Procyonidae, Carnivora) reveals extensive morphological convergence** ..... **97**
Introduction ..... 97
Materials and methods ..... 98
*Samples, amplification, and sequencing strategy* ..... 98
*Phylogenetic analyses* ..... 99
Results ..... 100
*General sequence results and insertion/deletion information* ..... 100
*Phylogeny of the Procyonidae* ..... 101
*Alternate topologies recovered for familial relationships within Musteloidea* ..... 102
Discussion ..... 102
*Phylogeny of the Procyonidae* ..... 102

*Alternate phylogenetic placements of the red panda* ..... 104
Conclusion ..... 105
References ..... 109

**5. Multiple individuals, multiple nuclear loci, and the importance of partitioning mitogenomic datasets in determining the phylogeny of true seals (Phocidae, Pinnipedia)** ..... **112**
Introduction ..... 112
Materials and methods ..... 115
*Sample selection, DNA isolation, amplification, and sequencing strategy* ..... 115
*Sequence alignment, model selection, base composition and marker congruence* ..... 117
*Tree search analyses* ..... 118
*Mitochondrial partitioning strategies* ..... 119
*Partition exclusion to avoid base composition bias across taxa and RY-coding* ..... 120
*Combined nuclear and mitochondrial analyses* ..... 120
*Topological effects of multiple individuals and inter- and intraspecific variation* ..... 120
Results ..... 121
*Sequencing results* ..... 121
*Base composition and congruence analyses* ..... 121
*Nuclear phylogenetic results* ..... 123
*Mitochondrial phylogenetic results* ..... 124
*MtDNA partitioning strategies* ..... 125
*Combined mt and nuclear results* ..... 126
*Use of multiple individuals per species* ..... 126
*Inter- and intraspecific variation* ..... 127
Discussion ..... 127
*Partitioning, model use, and conflicting dataset: How to choose?* ..... 127
*Higher-level phocid relationships* ..... 130
*Phocina species relationships* ..... 131
*Lobodontini species relationships* ..... 133
*Use of multiple individuals* ..... 134
*Species monophyly* ..... 134
Taxonomic implications and conclusions ..... 135
References ..... 165

**6. Multiple fossil calibrations and nuclear loci provide new insight into true seal biogeography and divergence times** ..... **173**

Introduction ..... 173

Methods ..... 175

Results and discussion ..... 178

*Nuclear vs. mitochondrial divergence time estimates and use of multiple fossil calibrations* ..... 178

*Origin of Pinnipedia, Otarioidea, and Phocidae* ..... 181

*Phocidae* ..... 183

*Phocinae: tribal origins and association with ice* ..... 184

*Phocini* ..... 186

*Early Monachinae and the Monachini* ..... 188

*Miroungini* ..... 189

*Lobodontini* ..... 190

Conclusions ..... 192

References ..... 201

**7. General discussion and conclusions** ..... **208**

Resolution for Carnivora: many genes, many analyses, many answers ..... 208

Themes in carnivore evolution ..... 210

Future directions ..... 211

References ..... 213

## List of Tables


| | | |
|---|---|---|
| Table 2-1 | Taxa included in study and accession numbers for sequences used for phylogenetic analysis | 39 |
| Table 2-2 | General sequence results by data partition | 43 |
| Table 2-3 | Maximum parsimony (MP) and maximum likelihood (ML) results by data partition | 44 |
| Table 3-1 | Classification of taxa included in study | 74 |
| Table 3-2 | Loci included in this study | 75 |
| Table 3-3 | Nuclear loci accession numbers and references | 77 |
| Table 3-4 | Mitochondrial gene accession numbers and references | 78 |
| Table 3-5 | Significant relative rate comparisons between lineages (GRate) | 79 |
| Table 3-6 | Locus information, ML results, and relative rate comparisons | 80 |
| Table 3-7 | Hypothesis tests | 81 |
| Table 3-8 | Effects of Mephitidae taxon sampling on the complete mtDNA result | 82 |
| Table 3-9 | Effects of outgroup selection on nuclear results | 83 |
| Table 4-1 | Species included in study | 106 |
| Table 4-2 | GenBank accession numbers for species included in study | 107 |
| Table 5-1 | Taxonomy of the Phocidae (and otarioid outgroups) | 137 |
| Table 5-2 | Individual phocid identification and collection information | 138 |
| Table 5-3 | GenBank accession numbers | 140 |
| Table 5-4 | Loci included in this study | 152 |
| Table 5-5 | Informative insertion-deletion information | 154 |
| Table 5-6 | Model information and maximum parsimony (MP) results | 155 |
| Table 5-7 | Clade support values for the Phocidae nuclear phylogeny | 156 |
| Table 5-8 | Clade support values for the Phocidae mitochondrial phylogeny | 157 |
| Table 6-1 | BEAST calibrations | 194 |
| Table 6-2 | Separate nuclear and mt BEAST divergence times | 195 |
| Table 6-3 | BEAST nuclear + mt divergence times | 196 |

## List of Figures


| | | |
|---|---|---|
| Figure 1-1 | Overview of the Arctoidea | 9 |
| Figure 2-1 | Family level phylogeny of the Caniformia | 45 |
| Figure 2-2 | Maximum likelihood tree recovered from the combined data set containing five nuclear genes (-*ln*L=18759.18170) | 46 |
| Figure 2-3a | Maximum likelihood cladogram for Ursidae recovered from all genes | 47 |
| Figure 2-3b | Maximum likelihood cladogram for Ursidae recovered from STS introns only | 47 |
| Figure 2-4 | Phylogeny of the Pinnipedia | 48 |
| Figure 2-5 | Phylogeny of the Mustelidae adapted from the maximum likelihood topology (Figure 2-1) | 49 |
| Figure 3-1 | Hypothesized relationships between musteloid families | 84 |
| Figure 3-2 | Saturation plots | 85 |
| Figure 3-3 | Maximum likelihood topology for mtDNA excluding 3$^{rd}$ codon positions | 86 |
| Figure 3-4 | Mt amino acid maximum likelihood topology | 87 |
| Figure 3-5 | Nuclear maximum likelihood topology | 88 |
| Figure 3-6 | NeighbourNet analysis of concatenated nuclear loci using LogDet distance | 89 |
| Figure 4-1 | Phylogeny of the Procyonidae | 108 |
| Figure 5-1 | Review of proposed phocine relationships | 158 |
| Figure 5-2 | Visualization of nuclear and mitochondrial base composition | 159 |
| Figure 5-3 | Nuclear phylogeny of the Phocidae | 160 |
| Figure 5-4 | Mitochondrial phylogeny of the Phocidae | 161 |
| Figure 5-5a | Phocina topology recovered by combined nuclear (no *GHR*) + mitochondrial (no 3$^{rd}$ positions) dataset | 162 |
| Figure 5-5b | Phocina topology recovered by combined nuclear (no *GHR*) + mitochondrial (by rate) dataset | 162 |
| Figure 5-5c | Lobodontini relationships in combined analyses | 162 |
| Figure 5-6 | Taxonomy versus mitochondrial genetic distance | 163 |
| Figure 5-7 | Taxonomy versus nuclear genetic distance | 164 |
| Figure 6-1 | Overview of pinniped relationships of extant and some extinct taxa. | 197 |

Figure 6-2 Combined nuclear + mt BEAST phylogeny..... 198
Figure 6-3 Hypothesized historical movements of phocid lineages..... 199
Figure 6-4 Present distribution of phocids..... 200

# Chapter 1
# General Introduction

## The Carnivora

The lives of carnivores (Carnivora, Mammalia) have long been intertwined with those of humans. We keep them as companions, hunt them for food and clothing, fear them as predators and respect them for their strength and beauty. But despite this close connection, we still know little about many aspects of their evolutionary history.

Carnivores show incredible diversity and are found in nearly every corner of the earth. Pinnipeds inhabit every ocean except the Indian, including many species living in either the arctic or Antarctic. Fissipeds (land-dwelling carnivores) are found on every continent except Antarctica, across an extreme range of habitat types and elevations (see Nowak, 2005). Consequently, carnivores exhibit a large range of morphological variation. Examples include wolf (*Canis lupus*) coat color ranging from white to black within a single pack and the extreme size difference between the smallest carnivore, the 25-250 g least weasel (*Mustela nivalis*) to the largest fissiped, the 300-800 kg polar bear (*Ursus maritimus*) and the largest pinniped, the ~900 kg male southern elephant seal (*Mirounga leonina*) (King, 1983; MacDonald and Kays, 2005; Nowak, 2005).

Behavioural variation among carnivores is equally great. Many species are highly social, living in small to large groups, while others are solitary, coming together only to mate (Gittleman et al., 2001). Carnivores range from predatory hunters (e.g. felids, hyaenids, canids, weasels), to omnivores (raccoons, bears), frugivores (kinkajou), and herbivores (giant and red pandas). Fecundity can also vary greatly. While most species bear between 1 and 6 offspring yearly, arctic foxes (*Alopex lagopus*) can have up to 25 young in a litter and a female mongoose (*Herpestes edwardsii*) has been observed to have five litters in 1.5 years (Nowak, 2005). Conversely, black bears (*Ursus maritimus*) and some large cats may not have any offspring for 3 or more consecutive years (Nowak, 2005).

Carnivores are often the top predator in an ecosystem, existing in comparatively low density and requiring large territory. As such, they are increasingly pushed out of their historical range as humans expand in range and density, reducing the availability of food and unfragmented habitat. Thirty-nine carnivores are IUCN-listed as endangered or highly endangered. Sixty more are listed at lower IUCN status levels, meaning that roughly one third of all carnivore species are of some conservation concern, including six of the world's eight bear species. The endangered

status of many carnivores is well recognized. The giant panda (*Ailuropoda melanoleuca*) is inextricably associated with the World Wildlife Federation; the polar bear (*Ursus maritimus*), whose southernly populations are diminishing with the progressive reduction in the number of sea ice days (Parks et al., 2006), is the poster-child of the effects of global warming. To preserve carnivore species for the future, meaningful and effective conservation plans must incorporate knowledge of their past. What are the conditions that led to their diversification and speciation? When do they thrive and when do they fail? A robust phylogeny provides an important framework within which to explore such questions.

## Carnivore systematics

### *Carnivoramorpha*

The order Carnivora belongs to the class Mammalia. The extant (presently living) Carnivora and their fossil relatives, the Viverravidae and the paraphyletic Miacidae, are together known as the Carnivoramorpha (Flynn and Wesley-Hunt, 2005; Wesley-Hunt and Flynn, 2005). Carnivoramorpha can be diagnosed using four dental characteristics, including the P4/m1 carnassial pair (Flynn and Wesley-Hunt, 2005). The P4/m1 carnassial pair of teeth are highly specialized for shearing in a scissor-like motion and represent the unique character that distinguishes Carnivoramorpha within Mammalia (Flynn and Wesley-Hunt, 2005). However, this trait is lost or severely reduced in many species that are no longer highly carnivorous.

### *Phylogeny and classification of extant Carnivora*

The order Carnivora is divided into two suborders, Feliformia (cat-like), and Caniformia (dog-like). Feliformia is comprised of the Felidae (cats), Viverridae (civets and genets), the monotypic *Nandinia* (African palm civet), Hyaenidae (hyaenas), Herpestidae (mongooses), and the Malagasy carnivorans (Flynn et al., 2005). Caniformia consists of four major lineages: Canidae (dogs and foxes), Musteloidea (weasels, raccoons, skunks, red panda), Pinnipedia (seals, fur seals, sea lions, walrus), and Ursidae (bears). The latter three lineages comprise the Arctoidea (bear-like carnivores) and are the focus of this thesis. A schematic overview of the Caniformia is presented in Figure 1-1.

When I began my research in 2003, several aspects of the arctoid phylogeny were poorly resolved (Figure 1-1). The affinity of the Pinnipedia to either Ursidae or Musteloidea has been examined extensively using both morphological and molecular evidence and both relationships have been proposed numerous times (for review, see Sato et al., 2006). Mitochondrial DNA cannot resolve this relationship, and though it is suggestive of a pinniped-ursid grouping, a polytomy is the best mitochondrial solution (Delisle and Strobeck, 2005). In Chapter 2, I approach this problem using nuclear DNA markers.

Musteloidea contains four families and is the most speciose of the three arctoid lineages (Figure 1-1). Ailuridae is represented by a single species, the red panda (*Ailurus fulgens*). Procyonidae is the raccoon family. The Mustelidae are the typical mustelids: weasels, badgers, otters, and their relatives. Relatively recently (Dragoo and Honeycutt, 1997), skunks were removed from their long-assumed position as the subfamily Mephitinae within Mustelidae and elevated to familial status as Mephitidae. The stink badger, *Mydaus*, was also removed from its previous subfamily Melinae (badgers) to become part of the new skunk family. Elevation of the subfamily Mephitinae to the family Mephitidae initiated a burst of research and hypotheses about familial relationships. Mitochondrial DNA phylogenies supported Mephitidae and Ailuridae as sister and Mustelidae (*sensu stricto*, excluding skunks) and Procyonidae as sister (Flynn et al., 2000; Delisle and Strobeck, 2005). This was a drastic change in thinking, as the traditional debate over the position of the red panda had been centered on whether the red panda was a member of the bear, 'panda', or raccoon families, or represented its own unique family. Investigation into the phylogenetic position of the red panda continues and is discussed in Chapters 2 and 3.

Relationships within musteloid families were equally unclear, but for different reasons. Procyonidae is comprised of five genera, most of which live in central and South America. Though most genera are unknown to the general public, the North American raccoon (*Procyon lotor*) is quickly recognized as an urban pest. The raccoon's ability to thrive in urban centres on anthropogenic food and habitat sources is admirable but economically problematic. However, we know almost nothing about the evolutionary history of the North American raccoon and its understudied and often endangered relatives. A single morphology-based phylogeny of the extant procyonid genera exists (Decker & Wozencraft 1991), and the first molecular-based phylogenies were not published until 2007. Published simultaneously with another study of similar findings (Koepfli et al., 2007), Chapter 4 represents the first molecular phylogeny of the extant procyonid genera and genus-level relationships appear resolved.

Conversely, relationships between the ~60 species of Mustelidae remain unresolved, despite extensive study. A comprehensive phylogenetic analysis of all mustelids is lacking, as species are found across most of the world, prohibiting, to date, even the most determined researchers from achieving complete sampling. Traditional taxonomy (Simpson, 1945; Wozencraft, 1993) recognized six subfamilies: Lutrinae (otters), Mustelinae (weasels, martens, fisher, and relatives), Melinae (badgers), Taxidiinae (American badger, *Taxidea taxus*), Mellivorinae (honey badger or ratel, *Mellivora capensis*), and the Mephitinae (skunks, now family Mephitidae). Relationships between these subfamilies remain somewhat uncertain, though recent studies support revision. I propose a new subfamily structure in Chapter 2.

Pinnipedia is comprised of three families, Phocidae (true seals), Otariidae (fur seals and sea lions), and Odobenidae (walrus), but has not always been considered monophyletic. First supported by the work of Mivart (1885), the idea of diphyly was revived in the 1960s and 1970s (McLaren, 1960b; Tedford, 1976; Repenning et al., 1979; de Muizon, 1982; Wozencraft, 1989b; Koretsky, 2001), supposing an ursid (bear) affinity for the Otariidae and Odobenidae and a mustelid (specifically otter) affinity for the Phocidae. The large body of evidence against diphyly includes historical classification (Flower and Lyddekker, 1891; Gregory and Hellman, 1939; Simpson, 1945), morphology (Wyss, 1987; Berta et al., 1989; Wyss, 1989; Wyss and Flynn, 1993; Berta and Wyss, 1994), and molecular studies comprising immunological (Sarich, 1969), karyotypic (Fay et al., 1967; Arnason, 1974, 1977), DNA-DNA hybridization (Arnason and Widegren, 1986), total evidence (Vrana et al., 1994; Bininda-Emonds et al., 1999) and sequence analyses (Vrana et al., 1994; Lento et al., 1995; Ledje and Arnason, 1996a, b; Flynn and Nedbal, 1998; Bininda-Emonds et al., 1999; Davis et al., 2004; Yu et al., 2004; Delisle and Strobeck, 2005; Arnason et al., 2006; Sato et al., 2006; Arnason et al., 2007; Higdon et al., 2007; Chapter 2). A monophyletic Pinnipedia is now universally accepted. However, the grouping of Otariidae and Odobenidae together in the superfamily Otarioidea has recently become contentious, as some prefer the phocid-odobenid grouping, Phocamorpha (Wyss, 1987, 1988; Berta and Wyss, 1994). The bulk of available evidence does not support Phocamorpha and Otarioidea continues to be supported and accepted by most researchers.

Rapid radiations within both Phocidae and Otariidae have led to difficulty in resolving many relationships (Wynen et al., 2001; Davis et al., 2004; Arnason et al., 2006; Higdon et al., 2007). In the otariids, this is exacerbated by reports of species and genus hybridizations (Goldsworthy et al., 1999; Brunner, 2002; Lancaster et al., 2006; Kingston and Gwilliam, 2007; Lancaster et al., 2007) that may compromise species delineation. Subfamilial and tribal

relationships within Phocidae are relatively well resolved. Phocidae is divided into two subfamilies: the northern seals, Phocinae, and the "southern" seals, Monachinae (the Antarctic seals and temperate-water monk and elephant seals). Relationships within the Antarctic tribe Lobodontini and the northern tribe Phocini remain unresolved, despite many morphological and several molecular investigations. Chapter 5 represents the largest complete molecular dataset used to address these relationships, presenting a new hypothesis for the Lobodontini and advances towards reconciling molecular hypotheses for the Phocini with morphology.

Biogeography has also been much discussed by pinniped researchers, particularly regarding the initial origin of Pinnipedia and the origin of the two land-locked species, the Caspian (*Pusa caspica)* and Baikal (*P. sibirica*) seals (Chapskii, 1955; Davies, 1958a, 1958b; McLaren, 1960a; Hendey, 1972; Repenning et al., 1979; de Muizon, 1982; Demere et al., 2003; Fyler et al., 2005; Arnason et al., 2006; Palo and Vainola, 2006). Biogeography builds upon a phylogeny to better understand the historical timing and environments that led to present diversity and distributions. Methods of incorporating fossil information and relaxing the molecular clock have drastically improved in recent years, allowing potentially more accurate estimation of the historical processes that lead to diversification. Molecular dating incorporating multiple fossil calibrations is employed in Chapter 6, in a discussion of phocid biogeography.

Comprised of only eight species, Ursidae is the smallest of the three arctoid lineages. Despite its small number of species, Ursidae can arguably be considered the most heavily studied arctoid group. Between 3 and 7 genera have been applied, though most recent classifications apply either 3 (Nowak, 2005) or 5 (Wozencraft, 1989a). The giant panda (*Ailuropoda melanoleuca*) is the only extant member of the subfamily Ailuropodinae, with all other species often grouped in the subfamily Ursinae. The spectacled bear (*Tremarctos ornatus*) is generally recovered as the first branch within the Ursinae and is often placed in its own subfamily, Tremarctinae. While now accepted, inclusion of the giant panda in Ursidae was historically contentious, as some authors favoured a grouping of the giant panda with the red, or lesser, panda (*Ailurus fulgens*). Characters grouping the two pandas are generally convergent based on their similar habitat and diet (Davis 1964, as quoted in Sarich, 1973) and the red panda is now accepted as a member of the Musteloidea. Relationships between the remaining six species are debatable and poorly resolved, excepting the close relationship between the polar (*Ursus maritimus*) and brown (*U. arctos*) bears (Talbot and Shields, 1996; Waits et al., 1999; Delisle and Strobeck, 2005). Until 2006, studies focusing on the bear family had either included large amounts of data, but not all species (Delisle and Strobeck, 2005), or very few genes for all species

(Talbot and Shields, 1996; Waits et al., 1999). In Chapter 2, I use increased data and include all species to provide a more comprehensive view of ursid relationships. Even more recent studies (Yu et al., 2007; Nakagome et al., 2008; Pages et al., 2008) have already expanded on this work to greatly improve our understanding of ursid evolution.

**Multiple gene analysis**

Recent developments in both sequencing technology and computing efficiency have led to tremendous increases in the amount of sequence data that can be analyzed for phylogenetic purposes. The field of phylogenomics is burgeoning (Murphy et al., 2004; Murphy et al., 2007; Nishihara et al., 2007), however, with these increases in data come numerous analytical challenges (Delsuc et al., 2005; Philippe et al., 2005a; Jeffroy et al., 2006; Nishihara et al., 2007; Rodriguez-Ezpeleta et al., 2007). Using multiple nuclear gene sequences means accounting for different rates of evolution across both genes and sites, changes in rate at a single site through time (heterotachy), differing levels of selection, nucleotide composition differences, aligning and modeling insertion-deletion events, gene duplications, and many more. Smaller datasets, such as those that I use to explore arctoid relationships, are vulnerable to the same problems.

Finding an appropriate means of combining data is not a new problem (Bull et al., 1993), nor are the problems of appropriately modeling DNA evolution. While much progress has been made in recent years, new problems are emerging. Applying a single evolutionary model to a concatenation of multiple genes has been shown to recover incorrect topologies; applying separate models to smaller partitions of the data can greatly improve phylogenetic accuracy (see Brown and Lemmon, 2007; Nishihara et al., 2007). Heterogeneity in evolutionary model fit within a gene can now be addressed using mixture models to allow different substitution models across sites (Lartillot and Phillipe, 2004; Venditti et al., 2008). Substitution rate heterogeneity across sites within a partition is often addressed using a discrete gamma approximation (Yang, 1994), although heterogeneity through time within a site (heterotachy) is an increasingly recognized problem (Lockhart et al., 2005; Philippe et al., 2005b; Zhou et al., 2007). Identifying base composition heterogeneity can help to avoid long branch attraction artifacts (see Lockhart et al., 1994; Galtier and Gouy, 1995; Jermiin et al., 2004). Distance-based models to better account for base composition biases across lineages have been available for some time (LogDet distance, Lockhart et al., 1994), but non-stationarity models for use in a likelihood context are more difficult to routinely implement at present (Yang and Roberts, 1995; Foster, 2004; Blanquart and

Lartillot, 2006; Boussau and Gouy, 2006). As more and more data become available, increasingly complicated patterns of DNA evolution are discovered and the suite of available analysis techniques to combat these problems in empirical datasets can be overwhelming.

How does one decide which evolutionary model is most appropriate for a given dataset? How many differently-evolving regions are identified and treated as independent partitions? With numerous analysis techniques often yielding different solutions, how does one choose which answer best reflects the true path of evolution? The many theoretical approaches to this diverse array of issues have been thoroughly tested using computer simulations, but it is sometimes unclear how they should best be applied to empirical data. Applying an inappropriate model of evolution can have serious consequences. If data violate the assumptions of the model, an incorrect topology may be recovered. At worst, inconsistency may occur, whereby increasing amounts of data lead to increasingly strong support for an incorrect topology (Felsenstein, 2004). As we increase the amount of sequence information under analysis, so too do we increase the heterogeneity in the dataset and the problem of appropriate modeling. Throughout this thesis, multiple analytical methods are applied to multi-gene datasets. This approach allows me to both better estimate the arctoid phylogeny and to explore and compare methods of identifying model violations and accounting for them when combining nuclear and mitochondrial datasets.

**Thesis objectives**

The objective of this thesis is two-fold. First, I aim to increase resolution of the arctoid phylogeny from species to superfamily levels. I approach several distinct phylogenetic problems within the Arctoidea using multiple DNA markers (Figure 1-1). Novel phylogenetic hypotheses are presented and evaluated with morphological and ecological consideration and taxonomic issues are addressed. Second, I address several issues that arise when combining multiple genetic markers, specifically those with differing inheritance patterns. These issues include: marker selection and data combinability; selecting between multiple approaches to partitioning and phylogenetic analysis; reconciling incongruence between mitochondrial and nuclear topologies; and identifying model violations.

More generally, it is my goal to provide analytical guidelines for similar phylogenetic studies and supply a strong framework for other studies of arctoid carnivores. Integrating physiological, behavioural, and ecological traits into an evolutionary framework allows identification of convergent versus ancestral adaptations, and an evolutionary timeframe can help

link past environmental changes with physical adaptations. Understanding historical responses of species and lineages to change can provide a strong basis for predicting their potential to adapt to ongoing environmental change. From an analytical perspective, examining ways of combining and analyzing multiple genes for phylogenetic purposes is broadly applicable. Many groups of researchers from across the world are collaborating to construct the Tree of Life (i.e. the Assembling the Tree of Life project, http://atol.sdsc.edu/). The Tree of Life is the 'ultimate' phylogeny containing all organisms, extant and extinct. Given the scope of such a project, experts in each section of the phylogeny contribute toward resolving their section of the phylogeny. This means that analyses are required at every level, from small subsections of the tree, such as the Carnivora, to the final, large, combined dataset. By providing guidelines and suggestions for resolving sections of the tree, I hope to contribute to other studies, beyond the Carnivora, bringing us closer, clade by clade, to estimating the entire Tree of Life. But while completion of the Tree of Life will represent a great advancement in knowledge, such a framework is a beginning, not an end goal. Considering all the genetic and environmental factors that lead to speciation and extinction bring us closer to understanding our present biological diversity and predicting, and protecting, its future.

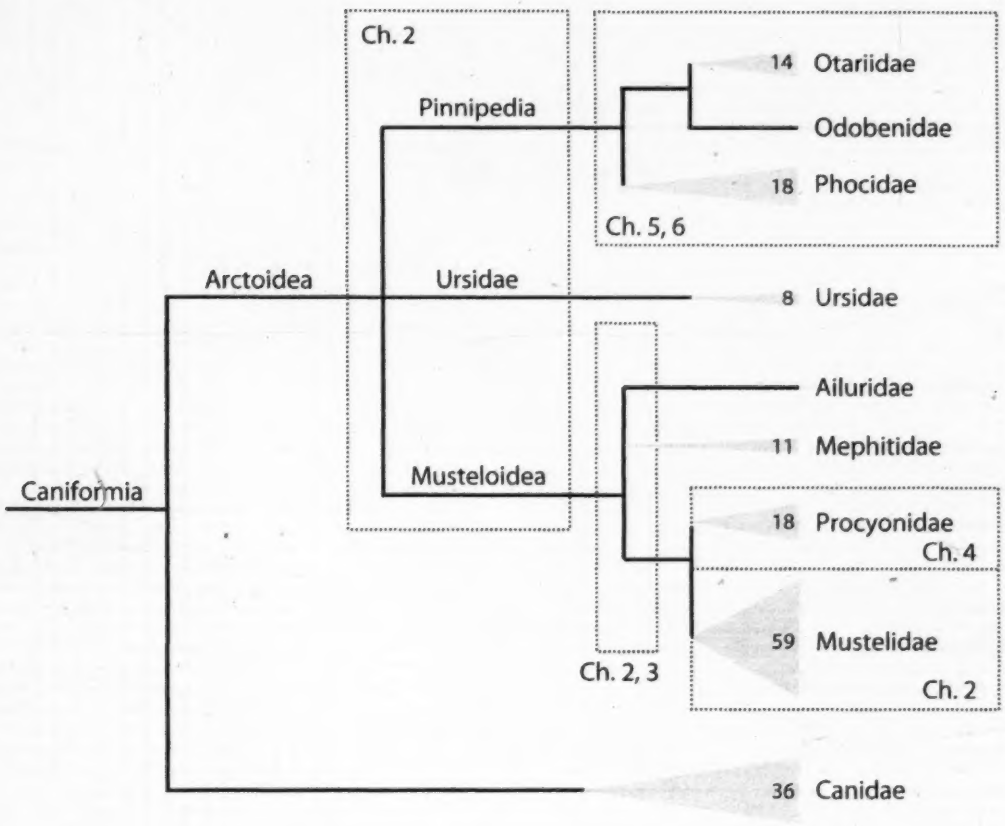


**Figure 1-1.** Overview of the Arctoidea. Dashed boxes represent the region of interest for the chapter indicated. The grey triangles represent each family, with the number of species indicated. Families with no grey box are represented by a single extant species.

**References**


Arnason, U., 1974. Comparative chromosome studies of Pinnipedia. Hereditas 76, 179-226.

Arnason, U., 1977. The relationship between the four principal pinniped karyotypes. Hereditas 87, 227-242.

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., 2007. Mitogenomic analyses of caniform relationships. Mol.Phylogenet.Evol. 45, 863-874.

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., Lehman, N., Petrov, E.A., Vainola, R., 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol.Phylogenet.Evol. 41, 345-354.

Arnason, U., Widegren, B., 1986. Pinniped phylogeny enlightened by molecular hybridizations using highly repetitive DNA. Mol.Biol.Evol. 3, 356-365.

Berta, A., Ray, C.E., Wyss, A.R., 1989. Skeleton of the oldest known pinniped, *Enaliarctos mealsi*. Science 244, 60-62.

Berta, A., Wyss, A.R., 1994. Pinniped phylogeny. Proc.San Diego Soc.Natl.Hist. 29, 33-56.

Bininda-Emonds, O.R.P., Gittleman, J.L., Purvis, A., 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biol.Rev. 74, 143-175.

Blanquart, S., Lartillot, N., 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogenous sequence evolution. Mol.Biol.Evol. 23, 2058-2071.

Boussau, B., Gouy, M., 2006. Efficient likelihood computations with nonreversible models of evolution. Syst.Biol. 55, 756-768.

Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst.Biol. 56, 643-655.

Brunner, S., 2002. A probable hybrid sea lion - *Zalophus californianus* X *Otaria byronia*. J.Mammal. 83, 135-144.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst.Biol. 42, 384-397.

Chapskii, K.K., 1955. Contribution to the problem of the history of development of the Caspian and Baikal seals. Trudy Zool. Inst. Akad. Nauk SSSR 17, 200-216.

Davies, J.L., 1958a. Pleistocene geography and the distribution of northern pinnipeds. Ecology 39, 97-113.

Davies, J.L., 1958b. The Pinnipedia: an essay in zoogeography. Geograph.Rev. 48, 474-493.

Davis, C.S., Delisle, I., Stirling, I., Siniff, D.B., Strobeck, C., 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. Mol.Phylogenet.Evol. 33, 363-377.

Davis, D.D., 1964. The giant panda: a morphological study of evolutionary mechanisms. Fieldiana Zool.Mem. 3, 1-339.

de Muizon, C., 1982. Phocid phylogeny and dispersal. Ann.S.Afr.Mus. 89, 175-213.

Decker, D. M., Wozencraft, W. C., 1991. Phylogenetic analysis of recent procyonid genera. J. Mammal. 72, 42-55.

Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Mol.Phylogenet.Evol. 37, 192-201.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat.Rev.Genet. 6, 361-375.

Demere, T.A., Berta, A., Adam, P.J., 2003. Pinnipedimorph evolutionary biology. Bull.Am.Mus.Nat.Hist. 13, 32-76.

Dragoo, J.W., Honeycutt, R.L., 1997. Systematics of Mustelid-like carnivores. J.Mammal. 78, 426-443.

Fay, F.H., Rausch, V.R., Feltz, E.T., 1967. Cytogenetic comparison of some pinnipeds (Mammalia: Eutheria). Can.J.Zool. 45, 773-778.

Felsenstein, J., 2004. Inferring phylogenies. Sinauer Associates, Inc, Sunderland, Massachusetts.

Flower, W.H., Lyddekker, R., 1891. An introduction to the study of mammals living and extinct. Adam & Charles Black, London.

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M.A., 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst.Biol. 54, 317-337.

Flynn, J.J., Nedbal, M.A., 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs incompatibility among multiple data sets. Mol.Phylogenet.Evol. 9, 414-426.

Flynn, J.J., Nedbal, M.A., Dragoo, J.W., Honeycutt, R.L., 2000. Whence the red panda? Mol.Phylogenet.Evol. 17, 190-199.

Flynn, J.J., Wesley-Hunt, G.D., 2005. Carnivora. In: Rose, K.D., Archibald, J.D. (Eds.), The rise of placental mammals: origins and relationships of the major extant clades. The John Hopkins University Press, Baltimore, Maryland, pp. 175-198.

Foster, P.G., 2004. Modeling compositional heterogeneity. Syst.Biol. 53, 485-495.

Fyler, C.A., Reeder, T.W., Berta, A., Antonelis, G., Aguilar, A., Androukaki, E., 2005. Historical biogeography and phylogeny of monachine seals (Pinnipedia: Phocidae) based on mitochondrial and nuclear DNA data. J.Biogeogr. 32, 1267-1279.

Galtier, N., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc.Natl.Acad.Sci. 92, 11317-11321.

Gittleman, J.L., Funk, S.M., MacDonald, D.W., Wayne, R.K., 2001. Why 'Carnivore Conservation'? In: Gittleman, J.L., Funk, S.M., MacDonald, D.W., Wayne, R.K. (Eds.), Carnivore Conservation. Cambridge University Press, Cambridge, UK, pp. 1-7.

Goldsworthy, S.D., Boness, D.J., Fleischer, R.C., 1999. Mate choice among sympatric fur seals: female preference for conphenotypic males. Behav.Ecol.Sociobiol. 45, 253-267.

Gregory, W.K., Hellman, M., 1939. On the evolution and major classification of the civets (Viverridae) and allied fossil and recent Carnivora: Phylogenetic study of the skull and dentition. Proc.Am.Phil.Soc. 81, 309-392.

Hendey, Q.B., 1972. The evolution and dispersal of the Monachinae (Mammalia: Pinnipedia). Ann.S.Afr.Mus. 59, 99-113.

Higdon, J.W., Bininda-Emonds, O.R.P., Beck, R.M.D., H., F.S., 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a mutigene dataset. BMC Evol.Biol. 7, 216.

Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? TRENDS Genet. 22, 225-231.

Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, A.W.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimates. Syst.Biol. 53, 638-643.

King, J.E., 1983. Seals of the world. British Museum (Natural History) and Oxford University Press, London and Oxford.

Kingston, J.J., Gwilliam, J., 2007. Hybridization between two sympatrically breeding species of fur seal at Iles Crozet revealed by genetic analysis. Conserv.Genet. 8, 1133-1145.

Koepfli, K.-P., Gompper, M.E., Eizirik, E., Ho, C.-C., Linden, L., Maldonado, J.E., Wayne, R.K., 2007. Phylogeny of the Procyonidae (Mammalia: Carnivora): molecules, morphology and the Great American Interchange. Mol.Phylogenet.Evol. 43, 1076-1095.

Koretsky, I.A., 2001. Morphology and systematics of Miocene Phocinae (Mammalia: Carnivora) from Paratethys and the North Atlantic region. Geologica Hungarica. Ser. Palaeo.Fasc. 54, 1-109.

Lancaster, M.L., Gemmell, N.J., Negro, S., Goldsworthy, S.D., Sunnucks, P., 2006. Ménage à trois on Macquarie Island: hybridization among three species of fur seal (*Arctocephalus* spp.) following historical population extinction. Mol.Ecol. 15, 3681-3692.

Lancaster, M.L., Goldsworthy, S.D., Sunnucks, P., 2007. Multiple mating strategies explain unexpected genetic mixing of New Zealand fur seals with two congenerics in a recently recolonized population. Mol.Ecol. 16, 5267-5276.

Lartillot, N., Phillipe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol.Biol.Evol. 21, 1095-1109.

Ledje, C., Arnason, U., 1996a. Phylogenetic analyses of complete cytochrome b genes of the order Carnivora with particular emphasis on the Caniformia. J.Mol.Evol. 42, 135-144.

Ledje, C., Arnason, U., 1996b. Phylogenetic relationships within caniform carnivores based on analyses of the mitochondrial 12S rRNA gene. J.Mol.Evol. 43, 641-649.

Lento, G.M., Hickson, R.E., Chambers, G.K., Penny, D., 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. Mol.Biol.Evol. 12, 28-52.

Lockhart, P.J., Novis, P., Milligan, B.G., Riden, J., Rambaut, A., Larkum, T., 2005. Heterotachy and tree building: a case study with plastids and eubacteria. Mol.Biol.Evol. 23, 40-45.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol.Biol.Evol. 11, 605-612.

MacDonald, D.W., Kays, R.W., 2005. Carnivores of the world: an introduction. In: Nowak, R.M. (Ed.), Walker's carnivores of the world. John Hopkins University Press, Baltimore, Maryland, pp. 1-67.

McLaren, I.A., 1960a. On the origin of the Caspian and Baikal seals and the paleoclimatological implication. Am.J.Sci. 258, 47-65.

McLaren, I.A., 1960b. Are the Pinnipedia biphyletic? Syst.Zool. 9, 18-28.

Mivart, S.G., 1885. Notes on the Pinnipedia. Proc. Zool. Soc. London 1885, 484-484-500.

Murphy, W.J., Pevzner, P.A., O'Brien, S.J., 2004. Mammalian phylogenomics comes of age. TRENDS Genet. 20, 631-632-639.

Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S., Miller, W., 2007. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. 17, 413-421.

Nakagome, S., Pecon-Slattery, J., Masuda, R., 2008. Unequal rates of Y chromosome gene divergence during speciation of the Family Ursidae. Mol.Biol.Evol. 25, 1344-1356.

Nishihara, H., Okada, N., Hasegawa, M., 2007. Rooting the eutherian tree: the power and pitfalls of phylogenomics. Genome Biol. 8, R199.

Nowak, R.M., 2005. Walker's carnivores of the world. John Hopkins University Press, Baltimore, Maryland.

Pages, M., Calvignac, S., Klein, C., Paris, M., Hughes, S., Hanni, C., 2008. Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. Mol.Phylogenet.Evol. 47, 73-83.

Palo, J.U., Vainola, R., 2006. The enigma of the landlocked Baikal and Caspian seals addressed through phylogeny of phocine mitochondrial sequences. Biol.J.Linn.Soc. 88, 61-72.

Parks, E.K., Derocher, A.E., Lunn, N.J., 2006. Seasonal and annual movement patterns of polar bears on the sea ice of Hudson Bay. Can.J.Zool. 84.

Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005a. Phylogenomics. Annu.Rev.Ecol.Evol.Syst. 36, 541-562.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., Delsuc, F., 2005b. Heterotachy and long-branch attraction in phylogenetics. BMC Evol.Biol. 5.

Repenning, C.A., Ray, C.E., Grigorescu, D., 1979. Pinniped biogeography. In: Boucot, A.J. (Ed.), Historical biogeography, plate tectonics, and thee changing environment: proceedings of the thirty-seventh annual biology colloquium and selected papers. Oregon State University Press, pp. 357-369.

Rodriguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst.Biol. 56, 389-399.

Sarich, V.M., 1969. Pinniped phylogeny. Syst.Zool. 18, 416-422.

Sarich, V.M., 1973. The giant panda is a bear. Nature 245, 218-220.

Sato, J.J., Wolsan, M., Suzuki, H., Hosoda, T., Yamaguchi, Y., Hiyama, K., Kobayashi, M., Minami, S., 2006. Evidence from nuclear DNA sequences sheds light on the phylogenetic relationships of pinnipedia: single origin with affinity to musteloidea. Zool.Sci. 23, 125-146.

Simpson, G.G., 1945. The principles of classification and a classification of mammals. Bull.Am.Mus.Nat.Hist. 85, 1-350.

Talbot, S.L., Shields, G.F., 1996. A phylogeny of the bears (Ursidae) inferred from complete sequences of three mitochondrial genes. Mol.Phylogenet.Evol. 5, 567-575.

Tedford, R.H., 1976. Relationship of pinnipeds to other carnivores (Mammalia). Syst.Zool. 25, 363-374.

Venditti, C., Meade, A., Pagel, M., 2008. Phylogenetic mixture models can reduce node-density artifacts. Syst.Biol. 57, 286-293.

Vrana, P.B., Milinkovitch, M.C., Powell, J.R., Wheeler, W.C., 1994. Higher level relationships of the arctoid Carnivora based on sequence data and "total evidence". Mol.Phylogenet.Evol. 3, 47-58.

Waits, L., Sullivan, J., O'Brien, S.J., Ward, R.H., 1999. Rapid radiation events in the family Ursidae indicated by likelihood phylogenetic estimation from multiple fragments of mtDNA. Mol.Phylogenet.Evol. 13, 82-92.

Wesley-Hunt, G.D., Flynn, J.J., 2005. Phylogeny of the Carnivora: basal relationships among the Carnivoramorphans, and assessment of the position of 'Miacoidea' relative to Carnivora. J.Syst.Palaeo. 3, 1-28.

Wozencraft, W.C., 1989a. Classification of the Recent Carnivora. In: Gittleman, J.L. (Ed.), Carnivore behaviour, ecology, and evolution. Cornell University Press, Ithaca, New York, pp. 569-593.

Wozencraft, W.C., 1989b. The phylogeny of the Recent Carnivora. In: Gittleman, J.L. (Ed.), Carnivore Behavior, Ecology, and Evolution. Cornell University Press, Ithaca, New York, pp. 495-535.

Wozencraft, W.C., 1993. Order Carnivora. In: Wilson, D.E., Reeder, D.A. (Eds.), Mammal species of the world: a taxonomic and geographic reference. Smithsonian Institution Press, Washington, DC, pp. 279-348.

Wynen, L.P., Goldsworthy, S.D., Insley, S.J., Adams, M., Bickham, J.W., Francis, J., Gallo, J.P., Hoetzel, A.R., Majluf, P., White, R.W.G., Slade, R.W., 2001. Phylogenetic relationships within the eared seals (Otariidae: Carnivora): Implications for the historical biogeography of the family. Mol.Phylogenet.Evol. 21, 270-284.

Wyss, A.R., 1987. The walrus auditory region and the monophyly of the pinnipeds. Am.Mus.Novit. 2871, 1-31.

Wyss, A.R., 1988. On "retrogression" in the evolution of the Phocinae and phylogenetic affinities of the monk seals. Am.Mus.Novit. 2924, 1-38.

Wyss, A.R., 1989. Flippers and pinniped phylogeny: has the problem of convergence been overrated? Marine Mamm.Sci. 5, 343-360.

Wyss, A.R., Flynn, J.J., 1993. A phylogenetic analysis and definition of the Carnivora. In: Szalay, F., Novacek, M., McKenna, M. (Eds.), Mammal Phylogeny: Placentals. Springer-Verlag, New York, pp. 32-52.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J.Mol.Evol. 39, 306-314.

Yang, Z., Roberts, D., 1995. On the use of nucleic acid sequences to infer early branchings in the tree of life. Mol.Biol.Evol. 12, 451-458.

Yu, L., Li, Q.-w., Ryder, O.A., Zhang, Y.P., 2004. Phylogenetic relationships within mammalian order Carnivora indicated by sequences of two nuclear DNA genes. Mol.Phylogenet.Evol. 33, 694-705.

Yu, L., Li, Y.-W., Ryder, O.A., Zhang, Y.-P., 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol.Biol. 7, 198.

Zhou, Y., Rodrigue, N., Lartillot, N., Philippe, H., 2007. Evaluation of the models handling heterotachy in phylogenetic inference. BMC Evol.Biol. 7, 206.

## Chapter 2

## Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets[1]

### Introduction

Monophyly of the order Carnivora (Mammalia) is well supported through multiple lines of evidence, including morphology (Flynn et al., 1988; Wyss and Flynn, 1993), DNA sequence data (Ledje and Arnason, 1996a, b; Flynn et al., 2000), total evidence (Vrana et al., 1994; Flynn and Nedbal, 1998), and supertree methods (Bininda-Emonds et al., 1999), as is the split between the two major suborders, Feliformia (cat-like carnivores) and Caniformia (dog-like carnivores). Within Caniformia, two infraorders exist: Cynoidea (or Canoidea), comprised of the single family Canidae (dogs and foxes), and Arctoidea, the bear-like carnivores. Arctoidea is further divided into three lineages of uncertain affinities: Ursidae (bears); Pinnipedia [families Phocidae (true seals), Otariidae (fur seals and sea lions), and Odobenidae (walrus)]; and Musteloidea [families Mustelidae (weasels and relatives), Procyonidae (raccoons), Mephitidae (skunks), and Ailuridae (red panda)].

The bear family, Ursidae, has been characterized by rapid radiation events, making phylogenetic inference of the species relationships difficult, and thus, often contentious (Waits et al., 1999; Yu et al., 2004a and references therein). Three subfamilies are generally recognized: Ailuropodinae (giant panda, *Ailuropoda melanoleuca*), Tremarctinae (spectacled bear, *Tremarctos ornatus*), and Ursinae (all other extant bears, genera *Ursus, Melursus,* and *Helarctos*). Ailuropodinae is now generally accepted as the earliest branching subfamily, though placement of the giant panda within the bear family (vs. its closest relative) has been a longstanding controversy (see both Waits et al., 1999; Yu et al., 2004a). Relationships among the ursine bears (sun, sloth, black, brown, and polar bears) remain relatively unresolved. The sloth bear (*Melursus ursinus*) is usually recovered as most basal, but the sun bear (*Helarctos malayanus*) has been placed as sister to the brown and polar bears (*Ursus arctos* and *U. maritimus*), the American and Asian black bears (*Ursus americanus* and *U. thibetanus*), or the sloth bear, though most of these placements have been poorly supported by bootstrapping (Talbot and Shields, 1996; Waits et al., 1999; Yu et al., 2004a; Delisle and Strobeck, 2005).

---

[1] A version of this paper has been published. Fulton, T.L. and C. Strobeck. 2006. Molecular Phylogenetics and Evolution 41, 165-181.

Within Pinnipedia, familial relationships have been contentious, but most current studies support Otariidae and Odobenidae (sometimes superfamily Otarioidea) as sister to Phocidae (Vrana et al., 1994; Ledje and Arnason, 1996a, b; Dragoo and Honeycutt, 1997; Flynn and Nedbal, 1998; Bininda-Emonds et al., 1999; Flynn et al., 2000; Davis et al., 2004; Flynn et al., 2005). The family Otariidae is comprised of two subfamilies: Otariinae (sea lions) and Arctocephalinae (fur seals). These subfamily designations may be inappropriate in light of evidence of various types suggesting that both Arctocephalinae and Otariinae may be paraphyletic (Bininda-Emonds et al., 1999; Brunner, 2000; Wynen et al., 2001; Davis et al., 2004; Delisle and Strobeck, 2005).

Within Musteloidea, familial relationships are more uncertain. Inclusion of the red panda, *Ailurus fulgens*, in Musteloidea is increasingly supported, but with uncertain affinity within the superfamily. Recent molecular studies including representative species from all caniform families place *Ailurus* either with Mephitidae as the sister clade to mustelids and procyonids using primarily mitochondrial DNA (Flynn et al., 2000; Delisle and Strobeck, 2005), or as the most basal musteloid lineage using nuclear and mitochondrial DNA (Flynn et al., 2005). Support for either hypothesis remains moderate to weak. At the subfamily level, Wozencraft (1993) closely followed Simpson (1945) in dividing Mustelidae into the subfamilies Mustelinae (weasels), Mephitinae (skunks), Melinae (badgers), Taxidiinae (American badger), Mellivorinae (honey badger), and Lutrinae (otters). Recent molecular evidence supports the elevation of skunks to family level (Vrana et al., 1994; Ledje and Arnason, 1996a; Dragoo and Honeycutt, 1997; Flynn et al., 2000; Domingo-Roura et al., 2005; Flynn et al., 2005, but see Sato, 2004), as the Procyonidae are recovered as sister to Mustelidae *sensu stricto* (excluding skunks), making the 'traditional' Mustelidae (including skunks) paraphyletic.

Using the most extensive taxon sampling of the Caniformia to date, this study addresses relationships ranging from species-level to superfamily-level, specifically (1) the relationships between Pinnipedia, Ursidae, and Musteloidea, (2) family relationships within Musteloidea, and (3) subfamily-level to species-level relationships within each of the three major arctoid lineages.

As sequencing becomes increasingly efficient via technological advances, molecular data is being quickly obtained on a large scale and is easily accessible through GenBank. This has great advantages in phylogenetics, as more complete taxon sampling within monophyletic lineages can lead to increased accuracy in the estimated phylogeny (Rannala et al., 1998). Conversely, as sequences are often obtained from different studies, sequence is often unavailable for the complete taxon set, leaving entire genes in one or more taxa to be considered as "missing"

in multiple gene analysis. Missing data itself may not be problematic for phylogenetic accuracy so long as enough informative characters are present to place the incomplete taxon within the phylogeny (Dragoo and Honeycutt, 1997; Wiens, 2003, 2006). Unfortunately, no *a priori* method exists to determine the number of required informative characters for complete resolution, as this is taxon and tree-dependent. If too few informative characters exist, 'floating' taxa may reduce overall support for clades that may otherwise be well supported. However, recent simulations suggest that the addition of taxa with up to 75% missing data may still not have a negative impact on the accuracy of the tree (Wiens, 2006). A more practical problem is the increase in tree search time associated with large amounts of missing data.

In the interest of including all taxa, even highly incomplete ones, in the final phylogeny, a considerable amount of the complete data set is missing. Four type I sequence-tagged sites (STS) and one nuclear exon, *IRBP*, including 190 newly generated sequences, were used in this study to address both the systematic questions for the Arctoidea described above and the effect that missing data may have on the final topologies from different analysis methods. Several methods of phylogenetic estimation were used: maximum parsimony, maximum likelihood, and Bayesian supermatrix analyses, and matrix representation with parsimony (MRP) supertree analysis. The first three methods are employed with incomplete taxa including missing data; the supertree method by definition avoids coding of missing data, as supertrees are constructed from source topologies, not sequence data. The MRP matrix is based on the nodes in each source tree using binary representation for the presence or absence of the taxa derived from each node (Baum and Ragan, 2004). Therefore, the matrix representation of individual topologies will include taxa as missing, but the individual gene source trees themselves will not contain missing data. Each individual gene data set contains between 57 and 71 of the total 85 included taxa, so all partitions are highly overlapping and thus, suitable for supertree methodology. MRP has been shown to be nearly as accurate as analysis of the combined primary data when the MRP input taxon sets are completely overlapping, though this accuracy decreases as the amount of taxon overlap decreases, analogous to an increase in the amount of missing data (Bininda-Emonds and Sanderson, 2001). Here, I investigate the empirical effect that this missing data may have on the supermatrix and supertree methodologies.

## Materials and methods

### *Samples, amplification, and sequencing strategy*

85 samples representing 79 caniforms and 6 feliform outgroup species were included in this study and are listed in Table 2-1. Sequences were obtained for four type I sequence-tagged sites (STS); (1) *feline sarcoma protooncogene (FES)* intron 14, (2) *cholinergic receptor, nicotinic, alpha 1 (CHRNA1)* intron 8, (3) *growth hormone receptor (GHR)* intron 9, (4) *rhodopsin (RHO)* intron 3; and one exon; *interphotoreceptor retinoid-binding protein (IRBP)* exon 1. Each region was PCR amplified and sequenced using amplification primers and additional internal primers, when necessary. Sources for previously available data from GenBank, including accession numbers, are listed in Table 2-1. New sequences have been entered into GenBank under the accession numbers DQ205725 to DQ205914.

Total genomic DNA was isolated from either tissue or blood using the QIAgen DNeasy Tissue Extraction kit. New sequences were generated using PCR amplification of each region, using ~150ng of template in a 50$\mu$l reaction containing 160$\mu$M dNTPs, 2.5mM $MgCl_2$, 1.5U of *Taq* DNA polymerase, and 1X PCR buffer (10mM Tris-HCl, pH 8.8, 0.1% Triton X-100, 50mM KCl, and 0.16 mg/mL bovine serum albumin). All PCR reactions were performed in a Perkin-Elmer 9600 thermal cycler; PCR primers and cycling conditions for the STS regions were as described in Koepfli and Wayne (2003): 3 minutes at 94°C, 30 cycles of 94°C for 30 sec, 54°C for 30 sec, 72°C for 45 sec, followed by a 5 minute hold at 72°C. Amplification of *IRBP* was performed as in Stanhope et al. (1992): 94°C for 5 min., followed by 35 cycles of 94°C for 1 min., 70°C for 3 min.. PCR products were resolved on a 1% agarose gel and isolated from excised gel fragments with the QIAquick gel extraction kit (Qiagen). Bi-directional sequencing was performed using BigDye v.3.1.1 (Applied Biosystems) following the manufacturer's protocol using the amplification primers. Internal sequencing primers were designed for *GHR* (*GHR*U.int GGAAAATTAGAAGAGGT, *GHR*L.int AAGAGTCATCGTTGTAGAA) and *IRBP* (+785 GGTACAGTGCCGACAAAGATG; -913 GCTTCTGGAGGTCCAGGGC) to ensure complete double-stranded sequencing. Sequences were resolved using an Applied Biosystems 3100 sequencer.

*Phylogenetic analysis*

Sequences were analyzed, basecalled, and aligned using 3100 Data Collection Software v.1.0.1 and Sequence Navigator v.1.0.1 (both from Applied Biosystems). Heterozygous sites (equal peak heights in electropherograms observed in both directions of sequence) were coded as polymorphisms. Alignments of newly generated sequences and those obtained from GenBank were performed both by eye and using the default settings in the EMBL-EBI online version of ClustalW (Chenna et al., 2003) to ensure consistency; no significant differences between methods were noted and final alignments were made by combining methods to achieve the minimal number of gaps (in frame, if applicable). Alignments have been entered into TreeBASE under the study accession number S1532 and matrix accession numbers M2755-6. All genes contained informative indels; these were coded as 0/1 (absence/presence of DNA, regardless of gap length) and included in maximum parsimony analyses with gaps in alignment delineated as "missing". Indel information was not included in maximum likelihood or Bayesian analyses, due to the DNA sequence evolution models employed. Sequences were partitioned into and analyzed by (1) individual genes, (2) all STS introns combined (i.e. excluding *IRBP*), and (3) all genes together. The partition containing only the STS introns was used to examine the effect of missing data on the maximum parsimony, maximum likelihood, and Bayesian methods, as there is considerably less missing data in this partition compared to the complete data set.

Prior to concatenation, partition homogeneity tests (ILD test, Farris et al., 1995) were performed in PAUP* v.4.10b (Swofford, 2003) to ensure congruence between all partitions. Twenty-four tests included: all data, all STS only, each gene to all remaining genes (5 tests), all pair-wise gene comparisons (10 tests), and, for *IRBP* alone: all coding positions, each coding position to the other two positions (3 tests), and all pair-wise coding positions comparisons (3 tests). All tests were non-significant at $\alpha$=0.05.

Maximum parsimony (MP) analyses were hindered by relatively long search times and large numbers of equally parsimonious trees due to the large number of missing characters and lack of sufficient informative characters, thus, limited search replicates were used. Heuristic searching was performed in PAUP* with TBR branch-swapping and as-is addition of taxa. Bootstrap support (BP) was obtained by 100 bootstrap replicates of heuristic searching with TBR branch-swapping and 10 random sequence addition replicates. Due to computer memory constraints, searches for individual genes were limited to hold 100 trees/addition sequence replicate; concatenated data sets were limited to 1000 trees/addition sequence replicate.

Modeltest v.3.0.6 (Posada and Crandall, 2001) was implemented for each gene, all STS genes, and all genes to determine the most appropriate model of DNA evolution based on hierarchical likelihood ratio testing. The selected models for each partition are shown in Table 2-3. An iterative approach to maximum likelihood (ML) heuristic searching (Sullivan et al., 2005, and references therein) in PAUP* was employed for each partition using the selected DNA evolution model from Modeltest, with the starting parameters estimated from the MP topology. Two rounds of searching with set parameters followed by estimation of parameters on the optimized topology resulted in stable parameter estimates and log likelihoods for all searches.

Bayesian analyses were performed using MrBayes v.3.1.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). For each individual gene partition, the DNA evolution model selected by ModelTest was implemented. For combined data sets, parameters were estimated separately for each gene partition based on the individual gene models. Four chains (using default temperature = 0.2 for the three heated chains) were run for 1 000 000 iterations, sampling the "cold" chain every 100 iterations. The first 10% of samples were discarded as "burn-in" after visualization in the program Tracer v.1.3 (Rambaut and Drummond, 2003). Two runs were performed for each data set to ensure proper convergence.

Three supertree analyses were performed: all individual gene maximum parsimony trees combined, all individual gene Bayesian trees combined, and all maximum likelihood trees combined. Hereon in, these supertrees will be respectively referred to as the MP-input, Bayes-input, and ML-input supertrees. The MP source trees were 50% majority-rule consensus trees; the Bayesian source trees were 90% majority-rule consensus trees of the Markov chain Monte Carlo (MCMC) sampled trees. The program RadCon (Thorley and Page, 2000) was used to create the MRP supertree matrix for each and the resultant matrices were analysed in PAUP* by maximum parsimony heuristic searching with random addition of taxa. All input trees were equally weighted.

To statistically assess the various hypotheses concerning the placement of the red panda (*Ailurus fulgens*) within Musteloidea, the nonparametric Shimodaira-Hasegawa test (SH test, Shimodaira and Hasegawa, 1999) was performed using all data. The SH test was implemented in PAUP* using 10 000 RELL approximation replicates, estimating parameters appropriate to the TVMef+I+$\Gamma$ likelihood model. Three alternate topologies were tested simultaneously: the maximum likelihood topology where Mephitidae (skunks) are the most basal musteloid lineage, the ML topology modified such that the red panda is the most basal musteloid lineage, and the ML topology modified to place the red panda as sister to the Mephitidae. Pairwise Templeton

tests (Templeton, 1983) were implemented in PAUP* for the purpose of a parsimony-based hypothesis test, comparing the two alternate topologies, in turn, to the ML topology.

## Results

### *General sequencing and tree estimation results*

ILD tests showed no significant incongruence between any of the partitions. All individual genes contained informative indels; the number of informative indels are shown in Table 2-2. All indels supported clades with high maximum parsimony boostrap (MP BP) support and did not contribute support to any weakly supported clades. The aligned length of each gene partition is shown in Table 2-2. A 204 base pair (bp) insertion in *GHR* in all canid species was removed from the alignment, as was a 219 bp insertion in *RHO* in the red panda. Sequence for *FES* for domestic cat, tiger, and bobcat was obtained but removed from the alignment due to difficulty in assessing homology (results not shown). In some cases, amplification products could not be obtained due to poor DNA quality or the presence of numerous non-specific amplification products under all PCR conditions tested. In these cases, sequence was not included in the final data matrix (Table 2-1).

For all partitions, multiple most parsimonious trees were obtained (Table 2-3). Both the *GHR* and complete data set MP searches were stopped due to computer memory constraints (92,600 MP trees). MP tree length, number of informative sites, consistency index (CI), and retention index (RI) are reported for each partition in Table 2-3. Individual gene topologies (MP BP) contained polytomies but did not differ significantly from each other or from the concatenated total data MP BP topology. Differences between trees were only found in areas of weak support; no "hard" incongruencies (opposing topologies supported by ≥80% bootstrap support) were observed.

All maximum likelihood searches yielded a single most likely tree (log likelihoods in Table 2-2), except *FES*, which recovered six equally likely trees. All *FES* trees were identical in topology and all parameters except the transition:transversion (Ti:Tv) ratio, Ti:Tv kappa, and gamma shape. However, the greatest difference in any one parameter estimate was only $2^{-4}$.

Multiple most parsimonious trees were found for all supertree analyses. A strict consensus was made for each search and all results are based upon these strict consensus trees.

*Relationships between Pinnipedia, Ursidae, and Musteloidea*

All phylogenies constructed from concatenated data sets (all STS and all data) yielded Arctoidea as monophyletic with Ursidae as the most basal arctoid lineage (Figure 2-1) with high support (98-99% MP BP; BPP=1.0). Individual gene trees were either unresolved between arctoid lineages, or had lower support (MP BP=67-93%). The only phylogenies that did not contain a monophyletic Arctoidea were the ML and Bayesian *IRBP* gene trees. These placed Canidae within the Arctoidea, as sister to Musteloidea, as was also found by Yu et al.'s (2004b) MP analysis (but not by ML or Bayesian analyses) of caniform *IRBP* sequences. However, the Canidae branch was approximately twice as long as the average branch from musteloid root to tip, indicating some discrepancy within the tree (results not shown). This result was not seen using the MP method, suggesting that long-branch attraction is not the cause of the unusual topology. Imposing a constraint for arctoid monophyly created a polytomy among the three arctoid lineages; all other relationships were unchanged and –lnL decreased from –6567.5753 to –6572.6696 (results not shown). When constrained out of Arctoidea in this fashion, the canids remained on an unusually long branch. Removal of canids from the *IRBP* data set did not change the remaining topology.

The ML phylogram (Figure 2-2) illustrates the close evolutionary relationships between lineages and between taxa, as indicated by the many short branches. Fewer than 0.005 substitutions/site occur between the branching of Ursidae before Pinnipedia and Musteloidea. Within lineages, the average branch length from root to tip is approximately 3-4 times greater within Musteloidea than any other major clade. This discrepancy in rate may be due to the reduced body size of musteloids compared to pinnipeds and ursids, and the many factors associated with body size such as metabolic rate or generation time.

*Species relationships within Ursidae*

Relationships within Ursidae were unresolved by both supermatrix and supertree analyses of the complete data set (Figure 2-3a). All methods recovered the giant panda, *Ailuropoda melanoleuca*, as the most basal lineage (68% MP BP, 0.98 BPP for all data), followed by the spectacled bear, *Tremarctos ornatus* (MP BP=80%, BPP=1.0). ML and the MP-input supertree recovered the sun bear, *Helarctos malayanus*, as branching next, followed by the sloth bear, *Melursus ursinus*. The MP-input supertree placed the sloth bear in a polytomy with the *Ursus* species. All other methods included the sun bear in this polytomy, only recovering the brown

bear, *Ursus arctos*, and polar bear, *Ursus maritimus*, as a clade. The STS-only data set contained only four ursid species (Figure 2-3b). The sloth bear was the most basal, followed by the sun bear, then the American black bear, *Ursus americanus*, and polar bear. The clade formed by the latter two was strongly supported (MP BP=95%, BPP=1.0) and consistent with the topology obtained from analyses of all genes combined. The placement of the sloth bear as basal to the other ursines was also well supported (MP BP=87%, BPP=0.94), but is inconsistent with the all-data topology. However, this is not a hard incongruency, as placement of the sun bear as more basal than the sloth bear in the all-data topology is not supported by either Bayesian or parsimony with bootstrapping. Both data sets recovered topologies consistent with present subfamilial classifications. The Ailuropodinae (giant panda, *Ailuropoda melanoleuca*) was most basal, then the Tremarctinae (spectacled bear, *Tremarctos ornatus*) and the Ursinae (all other ursids).

*Relationships within Pinnipedia*

Both concatenated data sets yielded pinniped monophyly (100% MP BP; BPP=1.0) in all construction methods, with strong support (100% MP BP; BPP=1.0) for the sister relationship of the Odobenidae (walrus) to the Otariidae (Figure 2-4). Below the family level, relationships are less well-supported. Within the Otariidae, the subfamily division between fur seals, Arctocephalinae, and sea lions, Otariinae, is not supported. Steller's sea lion, *Eumetopias jubatus*, is well supported (MP BP=94% for all-data, 83% for STS-only; BPP=1.0 for both data sets) as the most basal otariid included in this study; all other relationships were often polytomous depending on analysis method employed, though the South American fur seal, *Arctocephalus australis*, and New Zealand fur seal, *A. forsteri*, were moderately supported as sister taxa when all data were included (MP BP=76%; BPP=1.0). The positions of the Antarctic fur seal, *A. gazella*, and South American sea lion, *Otaria byronia*, differed depending on the construction method used. Both likelihood-based methods placed the two as a clade sister to the remaining fur seals (Figure 2-4), the ML-input supertree recovered the Antarctic fur seal as basal to the remaining otariids (Figure 2-4); all other methods did not resolve the relationship.

Relationships within the Phocidae are fairly well supported for higher-level designations, but often unresolved at the species level (Figure 2-4). Monophyly of the subfamily Monachinae (southern seals and monk seals) is strongly supported by all combined analyses (MP BP >90%; BPP=1.0). Remaining relationships are consistent with traditional tribal designations: the Monachini (monk seals) were the deepest branching lineage, followed by the northern elephant

seal (tribe: Miroungini) and the Lobodontini (all other southern seals). Relationships within the Lobodontini were unresolved by all methods except ML, which only recovers a single tree without support and the resolution may therefore not be significant. The exception to this is the sister relationship between the leopard, *Hydrurga leptonyx*, and Weddell, *Leptonychotes weddellii*, seals, found in all trees except the MP-input and ML-input supertrees (polytomies). Bayesian analysis weakly recovered (BPP=0.6 in both all-genes and STS-only data sets) the Ross seal, *Omatophoca rossii*, as the earliest branching lobodontine, followed by the crabeater seal, *Lobodon carcinophagus*. Monophyly of the subfamily Phocinae (northern seals) is moderately supported, with the bearded seal, *Erignathus barbatus*, as the earliest branching lineage. The Bayes-input supertree did not resolve the position of the bearded seal, placing it in a polytomy with the remaining northern seals and the southern seals. ML and Bayesian and the MP-input and Bayes-input supertree analyses of the complete combined data set supported the harp seal, *Pagophilus groenlandicus*, as the next branching taxon (MP BP<50%, BPP=1.0), while combined STS data and ML-input supertree supported the hooded seal, *Cystophora cristata*, in that position (MP BP= 71%; BPP=1.0). The ML-input supertree alternately placed the harp seal in a clade with the ringed seal, *Phoca hispida*. The ML topology placed the grey, *Halichoerus grypus*, and Caspian, *Phoca caspica*, seals as sister and the harbour, *Phoca vitulina*, and spotted, *Phoca largha*, seals as sister; these relationships were very weakly supported by Bayesian. All other species relationships remain unresolved.

*Relationships within Musteloidea*

Monophyly of the Musteloidea (including *Ailurus*) was strongly supported (MP BP=100%; BPP=1.0) by all data sets and all tree construction methods (Figure 2-1), as were the lineages corresponding to the family designations of Ailuridae (red panda), Mephitidae (skunks), Procyonidae (raccoons), and Mustelidae (weasels, badgers, and otters). Mephitidae is recovered as the earliest branching musteloid family, followed by Ailuridae, then Procyonidae and Mustelidae. Support for Procyonidae and Mustelidae as sister families is strong (MP BP=97% for total data, 94% for STS only; BPP=1.0). Placement of Mephitidae as more basal than Ailuridae is strongly supported by Bayesian analyses (BPP=1.0 for total data, BPP=0.98 for STS only), but only weakly supported by maximum parsimony (MP BP=59% for total data, <50% for STS only). This arrangement was also found in all strict consensus supertrees except the ML-input supertree, which recovered a trichotomy between Ailuridae, Mephitidae, and

Procyonidae+Mustelidae. The Shimodaira-Hasegawa (SH) test was implemented to test if the placement of Mephitidae as the most basal musteloid lineage is a statistically better arrangement than either the placement of Ailuridae as the most basal lineage or as sister to Mephitidae. The sister grouping of Ailuridae with Mephitidae was rejected at $\alpha$=0.05 and Ailuridae as the most basal musteloid lineage was rejected at $\alpha$=0.10 ($p$=0.092). The Templeton test, or Wilcoxon signed-ranks test (Templeton, 1983) was implemented under the parsimony criterion. Neither of the alternate topologies could be rejected at $\alpha$=0.05, consistent with the low parsimony bootstrap support (59%) recovered for the most parsimonious arrangement of Mephitidae as the most basal lineage.

The procyonid phylogeny is highly supported (MP BP=100%, BPP=1.0) at all nodes, using all data sets and construction methods (Figure 2-2). The kinkajou, *Potos flavus,* is basal to the other procyonids, consistent with the subfamily designations dividing the kinkajou and olingo (subfamily Potosinae) from the Procyoninae [raccoons, *Procyon*; ringtail, *Bassariscus*; and coatis, *Nasua*). However, an olingo (*Bassaricyon*) was not included in this study so the validity of these subfamily designations cannot be confirmed. The Procyoninae is also supported by one deletion event. Coatis are recovered as the most basal of the Procyoninae.

The mephitid phylogeny was also strongly supported (MP BP=100%, BPP=1.0) at every node, using all data sets and analysis methods. All three genera (*s.s.* excluding *Mydaus*) are represented, with the hog-nosed skunk, *Conepatus mesoleucus,* as basal to the western spotted skunk, *Spilogale gracilis*, and the striped skunk, *Mephitis mephitis* (Figure 2-2).

Within Mustelidae, major clades were well supported, but species relationships ranged from 100% support (MP BP and BPP) to a lack of resolution; some relationships differed between construction methods (Figure 2-5). In all combined analyses, the American badger, *Taxidea taxus*, is the earliest branch (all data and STS-only, respectively: MP BS=99%, 85%; BPP=0.99, 0.94), supporting its subfamilial status as Taxidiinae. Parsimony, Bayesian, and MP-input supertree analyses of both combined data sets weakly supported (MP BP=63% all-data, 88% STS-only; BPP=0.65 all-data, 0.99 STS-only) the Melinae (Old world badgers), excluding ferret-badgers (*Melogale*), as sister to the clade containing the genera *Martes*, *Eira,* and *Gulo* (Figure 2-5). Maximum likelihood did not place these two clades as sister, instead placing Melinae (excluding *Melogale*) as sister to all mustelines and lutrines. Bayes-input and ML-input supertree analyses could not resolve this relationship. Within the *Martes-Eira-Gulo* clade, the wolverine (*Gulo gulo*), tayra (*Eira barbara*), fisher (*Martes pennanti*), and yellow-throated marten (*Martes flavigula*) were basal to the remaining *Martes* species, but otherwise,

relationships were poorly resolved. The Chinese ferret badger, *Melogale moschata*, is strongly supported (MP BP=100%, BPP=1.0) as the basal lineage to the remaining mustelids, making Melinae polyphyletic. The remaining mustelids formed three lineages: the genus *Mustela*, the Lutrinae (otters), and the grison (*Galictis vittata*) + zorilla (*Ictonyx striatus*). All primary data supermatrix analyses placed *Mustela* as sister to a Lutrinae+(grison+zorilla) clade. Conversely, the ML-input supertree place the grison+zorilla clade as sister to *Mustela*, not Lutrinae. In all trees, the American mink, *Mustela vison*, and long-tailed weasel, *M. frenata*, formed a clade on the earliest branch in the *Mustela* clade, followed by the ermine, *M. ermina*, then the least and mountain weasels (*M. nivalis* and *M. altaica*) as a clade, and the European mink, *M. lutreola* as basal to the remaining *Mustela* species, which were found in a polytomy. The giant otter, *Pteronura brasiliensis*, was the most basal lutrine; all other otters were divided into two well-supported (MP BP>90%; BPP=1.0) clades. The *Lontra* clade is comprised of the river otter, *L. canadensis*, basal to the marine otter, *L. felina*, and neotropical otter, *L. longicaudis*. The second clade is composed of the sea otter, *Enhydra lutris*, and spotted-necked otter, *Lutra maculicollis*, on one branch, and the Eurasian otter, *Lutra lutra*, basal to the Oriental small-clawed and African clawless otters (*Amblonyx cinereus* and *Aonyx capensis*). The grouping of these genera is consistent with van Zyll de Jong's (1991) tribe Aonychini. Only the MP-input supertree method yielded a different lutrine topology, placing the sea otter and spotted-necked otters as branching sequentially instead of being sister taxa.

*Effect of missing data: all STS vs. total data*

To determine the effect of missing data on the different methodologies and their associated support values, tree searches were conducted for both the STS introns only and for the total data set (including *IRBP*). The total (all-data) data set has considerably more missing data than the STS-only data set, as 17 species only had sequence for *IRBP* and 10 species had sequence only for the four STS introns. The STS-only data set only contains 14 taxa that are incomplete to varying degrees (Table 2-1). The most significant difference between taxon sets is found within Ursidae and Mustelidae, where four and twelve taxa are added (respectively) to the STS-only taxon set to form the combined data set. The all-data analysis of Ursidae recovered the same topology (Figure 2-3a) as the *IRBP* gene tree, though all of the MP BP values decreased as much as 18% (*IRBP* gene tree results not shown). BPP values were essentially unchanged. No MP BP nodal support within Ursidae was higher than 85% in either tree, indicating that *IRBP*

does not contain enough information to clearly resolve the ursid topology, a problem only exacerbated by the addition of significant amounts of missing data in the complete data set. The addition of taxa within the Ursidae also had an effect. In the STS-only tree containing the polar, American black, sun, and sloth bears, the sloth bear was basal to the other bears. When all taxa were included in the analysis of all the data, the ML tree and the MP-based supertree recovered the sun bear as more basal. However, these results are not strongly incongruent, as placement of the sloth bear as basal by the STS-data set was strongly supported by both bootstrapping and Bayesian posterior probabilities, the ML all-data placement of the sun bear as more basal had no associated support. Within Mustelidae, addition of taxa with significant amounts of missing data (i.e. *IRBP* sequence only) did not conflict with the STS-only topology and a significant decrease in support was not observed, but most nodes in the all-data topology were not present in the STS-only topology because there were significantly fewer taxa included.

The converse situation occurred within the Lutrinae, where only two (the sea and river otters) of the nine included species have sequence for *IRBP*. Despite the lack of *IRBP* sequence for nearly all the lutrines, the topology and its Bayesian posterior probabilities remained unchanged between data sets and the MP bootstrap support was very similar. Of the eight nodes found within the Lutrinae, two had identical MP BP values in the STS and the total data trees, three had increased support with the total data set, and three had decreased support. However, the largest discrepancy between MP BP values was only 6%. Therefore, within the Lutrinae, the missing *IRBP* data appears to have had little to no effect, as the STS information was enough to resolve the topology. On the overall topology, the same is true. The two topologies are completely congruent, though the STS-only taxon set was smaller so several nodes cannot be observed compared to the total data tree. All nodes that were highly supported by MP BP in the STS data set were also highly supported in the total data set. Other less supported nodes had similar values between trees and no trend for either an increase or decrease in support between trees was observed. Only the Ursidae experienced a decrease in bootstrap support by the addition of a large amount of missing data.

*Supertree vs. supermatrix*

The three supertree topologies based on ML, MP, and Bayesian input trees were compared to the ML, MP BP, and Bayesian supermatrix topologies. In general, very few discrepancies were found between the supermatrix topologies and supertrees and most

discrepancies occurred in areas that were not well resolved by any method. In general, the differences between supertrees and supermatrix topologies were not greater than the differences found between the ML, Bayesian, and MP supermatrix methodologies. All trees resulted in identical family-level topologies, although the ML-input supertree did not resolve the relative position of the red panda within the Musteloidea. Five different species positions were obtained within the strict consensus supertrees that were not observed in the supermatrix trees. The MP-input supertree placed the sea otter and spotted-necked otters as branching sequentially (in that order), instead of forming a clade sister to the remaining species in the larger clade (Figure 2-5). The ML-input supertree exchanged the positions of the wolverine and yellow-throated marten within the *Martes-Eira-Gulo* clade and placed the grison+zorilla clade as sister to *Mustela* within Mustelidae (Figure 2-5). The ML-input supertree additionally placed the harp and ringed seals as sister within the Phocinae and did not group the Antarctic fur seal and South American sea lion as sister, instead, they branched sequentially (Figure 2-4). The Bayes-input supertree did not recover any clades that were not present in at least one of the supermatrix analyses. Although all of the supertrees resulted in many equally parsimonious trees, the strict consenses of these remain relatively well resolved, except in areas that remain poorly supported or unresolved by all methods.

## Discussion

### *Higher-level systematics of the Arctoidea*

Both data sets; all data combined and STS-only; and all analysis methods yielded Arctoidea as monophyletic and strongly supported Ursidae (MP BP=98%, BPP=1.0) as the most basal arctoid lineage (Figure 2-1). This arrangement for the Arctoidea was also strongly supported by the nuclear analysis of Flynn et al. (2005) and weakly supported by supertree analysis (Bininda-Emonds et al., 1999), but is not supported by mtDNA alone (Delisle and Strobeck, 2005), and is only sometimes supported by morphology (Wolsan, 1993; Wyss and Flynn, 1993). Within Arctoidea, monophyly of each family and major lineage (Ursidae, Pinnipedia, Musteloidea) is strongly supported (MP BP≥93%, BPP=1.0).

*Relationships within Ursidae*

All analyses of the total data set for the Ursidae supported the subfamilial designations of Ailuropodinae (giant panda, *Ailuropoda melanoleuca*), Tremarctinae (spectacled bear, *Tremarctos ornatus*), and Ursinae (all other ursids), with the latter two as sister (Figure 2-3a). Resolution within Ursinae was poor in the all-data analyses (Figure 2-3a), and though the sun bear, *Helarctos malayanus*, was recovered as more basal than the sloth bear, *Melursus ursinus*, this was a very short branch and was unsupported. Little confidence is placed in this resolution. The STS-only analyses containing only the sloth, sun, polar (*Ursus maritimus*), and American black bears (*Ursus americanus*) were strongly supported by both Bayesian and MP BP recovered the sloth bear, as the most basal ursine, and the sun bear, and the genus *Ursus* as a clade (Figure 2-3b). This placement of the sloth bear is congruent with previous molecular studies (Waits et al., 1999; Yu et al., 2004a), but strong support for the sun bear as more basal than both *Ursus* species is unusual, and contradictory to the mitochondrial DNA-based hypothesis placing the sun bear as sister to the brown and polar bears to the exclusion of the black bears (Waits et al., 1999; Delisle and Strobeck, 2005). This suggests that inclusion of the sun bear within the genus *Ursus* (vs. retaining *Helarctos*) as suggested by Nowak (1991), should not be implemented until a more well-supported consensus is reached.

*Relationships within Pinnipedia*

Within Pinnipedia, the Otariidae (fur seals and sea lions) and Odobenidae (walrus) were recovered as a sister clade to Phocidae (true seals), supporting the superfamily association of the former two as Otarioidea (Flynn and Wesley-Hunt, 2005). Otariidae is strongly supported as monophyletic, but the subfamilies Arctocephalinae (fur seals) and Otariinae (sea lions) are not. Steller's sea lion, *Eumetopias jubatus*, is strongly supported as the most basal otariid included, but other species relationships are not supported by all methods and are often unresolved (Figure 2-4). The placement of Steller's sea lion as more basal than the South American sea lion, *Otaria byronia*, conflicts with the results of Wynen et al. (2001), although their placement of *Otaria* as more basal was very weakly supported (MP BP<50%). Though the topologies differ, my results concur with that of the more comprehensive study of Wynen et al. (2001, and references therein), in that the present subfamily designations within Otariidae are misleading.

Phocidae is divided into two subfamilies, Phocinae (northern seals) and Monachinae (southern, elephant, and monk seals). This division is well supported by all methods (Figure 2-4), as are tribal distinctions to a lesser extent. However, at the species level, very little resolution exists and any resolved species relationships are generally associated with low support. Within the Phocinae, the bearded seal, *Erignathus barbatus,* is the only member of the tribe Erignathini and is well supported as the basal member of the subfamily. Relationships between and within the remaining two tribes, Cystophorini (monotypic hooded seal, *Cystophora cristata*) and Phocini (all other northern seals) are highly unresolved and differ based on the analysis method employed. Likelihood-based methods (ML and Bayesian) for the all-data set strongly supported the basal placement of the harp seal, *Pagophilus groenlandicus,* to *Cystophora*, making the Phocini paraphyletic. Within the remaining Phocini, the harbour and spotted seals (*Phoca vitulina* and *P. largha*) grouped as sister, as did the grey seal, *Halichoerus grypus*, and Caspian seal, *Pusa caspica*, but with poor support. In contrast, analyses of the STS-only data set, and the ML-input supertree strongly supported the hooded seal as branching after the bearded seal, consistent with mtDNA results (Davis et al., 2004) and supporting the tribal status of the hooded seal as Cystophorini and monophyly of the Phocini. Within Monachinae, the tribal relationships of Monachini (monk seals) as most basal, then Miroungini (elephant seals) and Lobodontini (southern seals) were supported (Figure 2-4). Within the Lobodontini, all methods except the ML-input supertree resolved the leopard seal, *Hydrurga leptonyx,* and Weddell seal, *Leptonychotes weddellii*, as sister. No other lobodontine species relationships were resolved, as is frequently the case within this tribe (Davis et al., 2004; Fyler et al., 2005).

*Musteloid family-level systematics: phylogenetic affinity of the red panda*

Within Musteloidea, monophyly is strongly supported (MP BP=100%, BPP=1.0) for each family: Mephitidae (skunks), Mustelidae (weasels, badgers, otters), and Procyonidae (raccoons). The family designation Ailuridae (red panda) is also supported. I propose a novel placement of the Ailuridae as sister to Mustelidae + Procyonidae, with Mephitidae as the most basal musteloid family (Figure 2-1). This topology was recovered by all analysis methods, except by the ML-input supertree, which could not resolve the relative position of the red panda. The basal placement of the Mephitidae has strong Bayesian support (1.0 for all genes, 0.98 for STS-only), but poor MP BP support (59% for all genes, <50% for STS-only). Flynn et al. (2005) obtained nearly identical support values (MP BP<50%, BPP=0.99-1.0) for their placement of the red panda

as the most basal musteloid, as did Delisle and Strobeck (2005) for their placement of the red panda as sister (MP BP=73-77%, BPP=1.0) to the Mephitidae (as proposed by Flynn et al., 2000). Likelihood-based statistical testing using the nonparametric Shimodaira-Hasegawa test of these alternate placements compared to the proposed placement rejected the first alternate hypothesis with $p<0.1$ and the second with $p<0.05$. The SH test is usually conservative, testing whether or not alternate hypotheses are equally good explanations of the data (Goldman et al., 2000). Because the test itself is so conservative, rejection at $\alpha=0.1$ is still a strong result. Parsimony-based hypothesis testing (Templeton test) could not reject either of the alternate topologies, consistent with the relative inability of parsimony to resolve this relationship. Because Ailuridae is monotypic and Mephitidae is on a relatively long branch, it is possible that long-branch attraction may be a factor in parsimony analysis. At present, I propose this novel placement of the red panda as a third hypothesis, but conservative interpretation would retain a polytomous relationship between Ailuridae, Mephitidae, and Mustelidae + Procyonidae. Based on available data, it appears that mitochondrial DNA analyses recover Mephitidae and Ailuridae as sister (Flynn et al., 2000; Delisle and Strobeck, 2005), nuclear DNA supports Ailuridae as branching after Mephitidae (this study), and a combination of the two yields a 'compromised' position of Ailuridae as the most basal musteloid lineage (Flynn et al., 2005). Future work combining morphology, nuclear DNA, mtDNA, and other molecular techniques will hopefully aid in the resolution of the phylogenetic affinity of the 'enigmatic' red panda.

*Species relationships within Mephitidae and Procyonidae*

The Mephitidae are comprised of four genera: *Spilogale*, spotted skunks; *Mephitis*, striped skunks; *Conepatus*, hog-nosed skunks; and *Mydaus*, the recently included stink badgers (Bryant et al., 1993; Dragoo and Honeycutt, 1997; Flynn et al., 2000). *Mydaus* was not represented in this study, but the remaining three genera were strongly supported as a monophyletic group, with *Conepatus* as most basal. This arrangement is congruent with the findings based on nuclear and mitochondrial DNA (Flynn et al., 2005), morphology (Bryant et al., 1993), and total evidence (Dragoo and Honeycutt, 1997).

Monophyly of Procyonidae is also strongly supported, as are all nodes within the family topology, using both data sets and all methods. I present the most comprehensive molecular phylogeny of the Procyonidae to date, including four of the five widely recognized genera. These are grouped into two subfamilies: Potosinae, represented by the kinkajou (*Potos flavus*) in this

study, and Procyoninae, which include the genera *Procyon* (raccoons), *Bassariscus* (ringtail), and *Nasua* (coatis). The fifth genus *Bassaricyon* (olingos), assigns to Potosinae, but it is not included in this study, precluding any definitive conclusions regarding subfamilial designations. The recovered topology (Figure 2-2) is consistent with the above subfamilial designations suggested by morphology (Decker and Wozencraft, 1991). The Procyoninae were recovered as monophyletic, with the clade supported by a single deletion event. Within Procyoninae, coatis are strongly supported as the most basal lineage, in contrast to morphology, which places the ringtail as more basal (Decker and Wozencraft, 1991). As more molecular data are obtained for *Bassaricyon*, both subfamilial designations and species relationships can be properly assessed.

*Subfamily and species relationships within Mustelidae*

Relationships within Mustelidae have been subject to considerable study both recently and in the past, with most recent studies in agreement upon the status (monophyletic or paraphyletic) of the subfamilies, though not necessarily on the species relationships within them (Bryant et al., 1993; Dragoo and Honeycutt, 1997; Koepfli and Wayne, 1998, 2003; Sato et al., 2003; Sato et al., 2004; Flynn et al., 2005). At the subfamily level for badgers, Taxiidinae (American badger, *Taxidea taxus*) is the most basal lineage and the traditional Melinae (Old World badgers) is polyphyletic if *Mydaus* (stink badgers) and *Melogale* (ferret-badgers) are included. If *Mydaus* and *Melogale* are excluded in accordance with recent studies, monophyly of the Melinae (*s.s.*) comprising of *Meles* (Old World badgers) and *Arctonyx* (hog badgers) is strongly supported. However, the placement of Melinae (*s.s.*) is uncertain. It either branches after *Taxidea* and before the remaining mustelids as found by ML, or as sister to the *Martes-Eira-Gulo* clade, as recovered and moderately supported by all other analysis methods (Figure 2-5). The Chinese ferret-badger, *Melogale moschata*, is strongly supported by all methods as the next branching lineage, further supporting its exclusion from Melinae. The remaining mustelids grouped in three strongly supported clades: *Mustela*, Lutrinae (otters), and the African zorilla (*Ictonyx striatus*) + the Central and South American-ranging grison (*Galictis vittata*). Most analysis methods placed the grison+zorilla clade as sister to Lutrinae, but the ML-input supertree instead recovered the clade as sister to *Mustela*. This sister relationship with *Mustela* was recovered by the individual gene trees for *CHRNA1*; all other individual gene trees did not resolve the relationship, indicating that the *CHRNA1* gene tree may have led to this relationship in the ML-input supertree. Monophyly of the Lutrinae was strongly supported (Figure 2-5). The

three major clades recovered within the Lutrinae correspond to the tribe Aonychini, the New World otters (*Lontra*), and the giant otter (*Pteroneura brasiliensis*). Most lutrine species relationships were well supported with one exception. The sea otter, *Enhydra lutris,* is recovered as either the most basal lineage in the Aonychini (*Enhydra-Lutra-Amblyonyx-Aonyx*) by the MP-input supertree or as sister to the spotted-necked otter, *Lutra maculicollis* by all other methods. The former arrangement was recovered by Koepfli and Wayne (2003) in their combined analysis of five nuclear STS loci (four of which are used in this study), but when only the STS loci were analyzed, the latter was found. However, though this sister relationship was recovered most often, it is essentially unsupported (MP BP=52% for both all-data and STS-only; BPP=0.69 for all-data and 0.64 for STS-only) and should be interpreted with caution. More in-depth systematics of the Lutrinae are discussed by Koepfli and Wayne (1998). Results within *Mustela* were the same as recovered by the MP analysis of Sato et al. (2003), though with somewhat less resolution as some relationships within *Mustela* are solely based on the *IRBP* gene tree. Species relationships are poorly resolved within the *Martes-Eira-Gulo* clade (Figure 2-5), though the clade itself is well supported. Four taxa, the fisher (*Martes pennanti*), tayra (*Eira barbara*), wolverine (*Gulo gulo*), and yellow-throated marten (*Martes flavigula*), are more basal on the tree than the remaining *Martes* species, but the precise relationships of these taxa is uncertain. However, the well-supported inclusion of the tayra and wolverine in this clade make the genus *Martes* paraphyletic. As a result, it is perhaps misleading for the fisher and/or the yellow-throated marten to be included in *Martes*, or for *Eira* and *Gulo* to be retained as distinct genera. Though genus-level classifications remain contentious, given the strongly supported monophyly of this group and the subfamilial status of the Melinae (*s.s.*) as the potential sister clade to the *Martes-Eira-Gulo* clade, I propose that this clade also be afforded subfamilial status as Martinae. Accordingly, if Mustelinae are restricted to the genus *Mustela*; the subfamily Galictinae (see Anderson, 1989) is reinstated to include *Galictis* and *Ictonyx* but not *Eira*; and the ferret-badgers (*Melogale*) become the subfamily Helictidinae as discussed by Sato et al. (2004), the subfamilies Taxidiinae, Melinae (*s.s*), and Lutrinae would remain valid. The mustelid genera *Vormela*, *Lyncodon*, *Poecilictis*, *Poecilogale*, and *Mellivora* are not included in this study and further work is necessary to confirm the phylogenetic affinities of these genera. *Mellivora* (honey badger) is generally considered as the monotypic subfamily Mellivorinae. *Poecilogale* has been closely allied with *Ictonyx* (Flynn et al., 2005) and *Poecilictis* may not even be a distinct genus from *Ictonyx,* thus, both will likely be included in the redefined Galictinae. Very little molecular work has been performed on *Vormela* and *Lyncodon,* but *Lyncodon* was originally included in

Galictinae (Anderson, 1989) and *Vormela* has also been allied with *Ictonyx* (Bininda-Emonds et al., 1999). However, definitive classification of these relatively unstudied species should be delayed until more information is acquired.

*Effect of missing data: all STS vs. total data*

In addition to systematic questions within the Arctoidea, I was also interested in the effect of missing data on the final topologies and their associated measures of support. It is sometimes the case that taxa missing data for entire genes are excluded from analysis to avoid the potential problems of reduced support and resolution (Flynn et al., 2005). However, increased taxon sampling within lineages, especially those taxa whose placement will break up long branches (thereby reducing long-branch attraction), has been shown to increase phylogenetic accuracy, more so than the addition of characters (Graybeal, 1998; Rannala et al., 1998). It has also been proposed that missing data itself, or even the proportion of missing data may not be problematic, so long as enough informative characters are present to place the taxon on the tree (Wiens, 2003). Inclusion of even highly incomplete taxa (10-25% complete) using model-based analyses such as ML or Bayesian or moderately incomplete taxa (50%) with MP analyses has been shown to have a strong positive effect on accuracy in simulation studies (Wiens, 2006). Based on this, it was my intention to include as many taxa as possible within the Arctoidea for which information for at least one of the included genes was available. To address the effects of missing data, two data sets were analyzed: one containing all five genes and 79 taxa, and one containing only the four STS loci for 62 taxa. With the addition of the *IRBP* gene to the STS data set, 17 taxa containing only sequence for *IRBP* were added, introducing a considerable number of missing data cells into the data set, especially within Ursidae and Mustelidae (Table 2-1). The result of the addition of these taxa (discussed below) and the associated missing data supported the idea (Wiens, 2003, 2006) that additional taxa have a greater impact on topology than additional characters and that the amount of missing data is less important than the number of informative characters for those taxa. Within Ursidae, four of the eight taxa were included with only *IRBP* sequence, resulting in a topology that matched that of the *IRBP* gene tree. The addition of taxa (and thus, missing data for all STS introns in those species) led to a general reduction in MP bootstrap support from the *IRBP* gene tree to the combined data set. However, all MP BP values were moderate to low and the largest decrease was 18%. BPP values were essentially unchanged between the *IRBP* gene tree and the all-data tree, indicating that Bayesian analyses are

more robust when faced with substantial missing data (~60% missing), likely due to lower susceptibility to long-branch attraction than MP (Huelsenbeck, 1995; Alfaro et al., 2003). Within Mustelidae, taxa with only *IRBP* sequence (~40% missing) were mainly added within *Martes* and *Mustela*. Within both genera, the STS-only and combined data topologies were completely congruent and support was not substantially reduced. Species relationships that could not be resolved in the *IRBP* gene tree remained unresolved in the combined tree, as no additional informative characters were added to successfully resolve relationships. Conversely, within Lutrinae (otters), the addition of *IRBP* added sequence for only two of the nine species, introducing missing data without much benefit of additional information. In this case, the topology was completely resolved by the STS-only sequence information and the addition of missing data had no clear effect. Equal numbers of nodes experienced increased MP BP support to those that decreased in support and the changes were small. In all cases where support was 100% MP BP or BPP=1.0, it remained unchanged, indicating that enough informative characters were present to nullify any effect of missing data. Overall, topologies that were well resolved prior to the incorporation of missing data remained well-resolved, while poorly resolved topologies experienced some decreased MP BP support. Lack of resolution in the supertrees appears to be due to the absence of enough informative characters to fully resolve input trees. Although not conclusive, these findings support the idea that if enough informative characters exist to place a taxon within a topology, the addition of missing data (to an extent) will have little to no effect.

*Supertree vs. supermatrix*

A second approach used to deal with the amount of missing data in the complete data set was the application of the matrix representation with parsimony (MRP) supertree method. Due to non-identical taxon sets between genes, a traditional consensus of the gene trees could not be made. Supertree methodologies are used to combine topologies, but do not require identical taxon sets, merely overlapping ones, as its construction algorithm is node-based, not tree-based (Baum and Ragan, 2004). The supertree method provided two major advantages to the analysis in this study. First, compared to a supermatrix approach, it is a considerably faster method for creating a ML-based tree. The creation of the supermatrix ML tree took approximately twice as long as ML searches for individual gene trees and subsequent construction of the ML-input supertree. Second, it allowed the gene information interpreted under MP, ML, and Bayesian

criteria to be combined in a different way for comparison with the supermatrix-derived topologies from these methods to examine the potential effect that missing data was having on these topologies.

The MRP method and its applications have been heavily criticized since it was proposed independently by two researchers (Baum 1992; Ragan, 1992). I do not address any methodological shortcomings that have been raised against the MRP method, but do believe that the method employed here avoids many of the shortcomings in previous applications of the MRP method. Combination of newly constructed gene trees via supertree analysis has been proposed in the past (Bininda-Emonds, 2004), avoiding problems such as non-independence (i.e. pseudoreplication) of source-tree data and questionable quality of source trees.

In this study, supermatrix and supertree topologies were highly congruent. Five branching patterns were recovered by only supertree methods and not by supermatrix analyses, and one of these "novel" arrangements, the sequential branching of the sea otter and spotted-necked otter (vs. sister), is actually the preferred topology by another molecular study (Koepfli and Wayne, 2003). The other four novel groupings were found in the ML-input supertree, likely reflecting the more resolved, but unsupported input topologies rather than truly novel clades. All supertree and supermatrix methods recovered the same subfamily and higher topologies, excepting the different placement of Melinae (*s.s.*) found only by ML supermatrix analysis (Figure 2-5). Conflicts between analytical methods and novel supertree clades were confined to areas of weak topological support, generally species relationships within subfamilies. In this study, differences appear to result from the need for increased data, not from specific shortcomings of any method (Bininda-Emonds, 2004).

Differences between ML and Bayesian methods are thought to have arisen from the ability to apply and estimate parameters for distinct data partitions (genes) within MrBayes for Bayesian analysis, and not within PAUP* for ML, as well as the fact that ML nearly always returns a single resolved topology as it does not rely on any type of consensus. If it were feasible to perform bootstrapping on our ML data set, I would have been able to identify areas that, while resolved, were weakly supported and were perhaps over-contributing to the ML-input supertree. All of the MP searches of the MRP-constructed matrices recovered thousands of equally parsimonious trees. The 50% majority-rule consensus of each of these searches returned a nearly completely resolved topology (results not shown). However, almost all of the nodes that were not found in 100% of the MP trees were found in less than 75%. I therefore used only the strict consensus of these trees (vs. majority-rule), in favour of obtaining a single supertree at the

expense of some (potentially misleading) resolution (Steel et al., 2000). Overall, when the resultant supertrees were compared to their supermatrix counterparts, very few topological differences existed, and both methods appeared to suffer equally from inadequacies in the primary data (too few informative characters). Based on this, the ML-input supertree analysis is advantageous to a ML supermatrix analysis, based on the shorter tree construction time and its more consensus-like basis when support cannot be obtained for the ML topology. However, the Bayesian supermatrix analysis of the combined data set was faster than the creation of individual Bayesian gene trees and their subsequent supertree construction. Bayesian analysis in MrBayes also has the benefit of applying gene-specific models in combined gene analysis and has an associated support measure. However, based on the results of this study, I advocate the use of several analysis methods for any data set, particularly those lacking enough informative characters to fully resolve relationships.

**Table 2-1.** Taxa included in study and accession numbers for sequences used for phylogenetic analysis. Blank cells indicate that sequence was unavailable due to either the absence of sequence in GenBank or to unsuccessful amplification or sequence. References for sequences obtained from GenBank are noted by: [a]Yu et al. (2004) [b]Sato et al. (2003) [c]Koepfli and Wayne (2003) [d]Sato et al. (2004) [e]Stanhope et al. (1992)
All other sequences were generated by this study.

| Family | Species | *IRBP* Exon I | *FES* | *GHR* | *CHRNA1* | *RHO* |
|---|---|---|---|---|---|---|
| Ursidae | *Ursus arctos* Brown bear | AY303842[a] | | | | |
| | *Ursus maritimus* Polar bear | AY303843[a] | DQ205765 | DQ205798 | DQ205725 | DQ205839 |
| | *Ursus americanus* American black bear | AY303837[a] | DQ205766 | DQ205799 | DQ205726 | DQ205840 |
| | *Ursus thibetanus* Asian black bear | AY303841[a] | | | | |
| | *Helarctos malayanus* Sun bear | AY303839[a] | DQ205767 | DQ205800 | DQ205727 | DQ205841 |
| | *Melursus ursinus* Sloth bear | AY303838[a] | DQ205768 | DQ205801 | DQ205728 | DQ205842 |
| | *Ailuropoda melanoleuca* Giant panda | AY303836[a] | | | | |
| | *Tremarctos ornatus* Spectacled bear | AY303840[a] | | | | |
| Procyonidae | *Procyon lotor* Raccoon | AB082981[b] | AF498183[c] | AF498207[c] | AF498152[c] | AF498231[c] |
| | *Procyon cancrivorous* Crab-eating raccoon | AB109332[d] | | | | |
| | *Nasua narica* White-nosed coati | DQ205878 | DQ205769 | DQ205802 | DQ205729 | DQ205843 |
| | *Nasua nasua* Coati | AY525031[a] | | | | |
| | *Bassariscus astutus* Ringtail | DQ205879 | AF498182[c] | AF498206[c] | DQ205730 | DQ205844 |
| | *Potos flavus* Kinkajou | DQ205880 | DQ205770 | DQ205803 | DQ205731 | DQ205845 |
| Ailuridae | *Ailurus fulgens* Red panda | DQ205881 | DQ205771 | DQ205804 | DQ205732 | DQ205877 |
| Mephitidae | *Mephitis mephitis* Striped skunk | AB109331[d] | | DQ205805 | DQ205733 | DQ205846 |
| | *Spilogale gracilis latifrons* Western spotted skunk | DQ205882 | DQ205772 | DQ205806 | DQ205734 | DQ205847 |
| | *Conepatus mesoleucus* Hog-nosed skunk | DQ205883 | DQ205773 | DQ205807 | DQ205735 | DQ205848 |
| Mustelidae | *Enhydra lutris* Sea otter | AB082978[b] | AF498162[c] | AF498186[c] | AF498131[c] | AF498210[c] |
| | *Aonyx capensis* African clawless otter | | AF498160[c] | AF498184[c] | AF498129[c] | AF498208[c] |
| | *Amblonyx cinereus* Oriental small-clawed otter | | AF498161[c] | AF498185[c] | AF498130[c] | AF498209[c] |
| | *Lontra canadensis* River otter | DQ205884 | AF498163[c] | AF498187[c] | AF498132[c] | AF498211[c] |

**Table 2-1.** *Continued.*

| Family | Species | *IRBP* Exon I | *FES* | *GHR* | *CHRNA1* | *RHO* |
|---|---|---|---|---|---|---|
| Mustelidae continued | *Lontra felina* Marine otter | | AF498164[c] | AF498188[c] | AF498133[c] | AF498212[c] |
| | *Lontra longicaudis* Neotropical otter | | AF498165[c] | AF498189[c] | AF498134[c] | AF498213[c] |
| | *Lutra lutra* Eurasian otter | | AF498166[c] | AF498190[c] | AF498135[c] | AF498214[c] |
| | *Lutra maculicollis* Spotted-necked otter | | AF498167[c] | AF498191[c] | AF498136[c] | AF498215[c] |
| | *Pteronura brasiliensis* Giant otter | | AF498168[c] | AF498192[c] | AF498137[c] | AF498216[c] |
| | *Taxidea taxus* American badger | DQ205885 | AF498179[c] | AF498203[c] | AF498148[c] | AF498227[c] |
| | *Meles meles meles* European badger | AB082980[b] | AF498178[c] | AF498202[c] | AF498147[c] | AF498226[c] |
| | *Meles meles anakuma* Japanese badger | AB082979[b] | | | | |
| | *Arctonyx collaris* Hog badger | AY525049[a] | AF498180[c] | AF498204[c] | AF498149[c] | AF498228[c] |
| | *Melogale moschata* Chinese ferret badger | AB109330[d] | AF498181[c] | AF498205[c] | AF498150[c] | AF498229[c] |
| | *Martes pennanti* Fisher | | AF498173[c] | AF498197[c] | AF498142[c] | AF498221[c] |
| | *Martes americana* American marten | AB082963[b] | AF498172[c] | AF498196[c] | AF498141[c] | AF498220[c] |
| | *Martes flavigula* Yellow-throated marten | AB082964[b] | | | | |
| | *Martes zibellina* Sable | AB109329[d] | | | | |
| | *Martes foina* Beech marten | AB082965[b] | | | | |
| | *Martes melampus* Japanese marten | AB082967[b] | | | | |
| | *Martes martes* Pine marten | AB082966[b] | | | | |
| | *Gulo gulo* Wolverine | AB082962[b] | AF498174[c] | AF498198[c] | AF498143[c] | AF498222[c] |
| | *Mustela vison* American mink | AB082977[b] | AF498171[c] | AF498195[c] | AF498140[c] | AF498219[c] |
| | *Mustela putorius furo* Domestic ferret | AB082974[b] | DQ205774 | DQ205808 | DQ205736 | DQ205849 |
| | *Mustela ermina* Ermine | AB082969[b] | AF498169[c] | AF498193[c] | AF498138[c] | AF498217[c] |
| | *Mustela frenata* Long-tailed weasel | DQ205886 | AF498170[c] | AF498194[c] | AF498139[c] | AF498218[c] |
| | *Mustela nivalis* Least weasel | AB082973[b] | DQ205775 | DQ205809 | DQ205737 | DQ205850 |
| | *Mustela lutreola* European mink | AB082972[b] | | | | |
| | *Mustela putorius* European polecat | AB082975[b] | | | | |

**Table 2-1.** *Continued.*

| Family | Species | *IRBP* Exon 1 | *FES* | *GHR* | *CHRNA1* | *RHO* |
|---|---|---|---|---|---|---|
| | *Mustela sibirica* Siberian weasel | AB082976[b] | | | | |
| | *Mustela eversmannii* Steppe polecat | AB082970[b] | | | | |
| | *Mustela itatsi* Japanese weasel | AB082971[b] | | | | |
| | *Mustela altaica* Mountain weasel | AB082968[b] | | | | |
| | *Eira barbara* Tayra | | AF498175[c] | AF498199[c] | AF498144[c] | AF498223[c] |
| | *Galictis vittata* Greater grison | | AF498176[c] | AF498200[c] | AF498145[c] | AF498224[c] |
| | *Ictonyx striatus* Zorilla | | AF498177[c] | AF498201[c] | AF498146[c] | AF498225[c] |
| Otariidae | *Arctocephalus australis* South American fur seal | DQ205887 | DQ205776 | DQ205810 | DQ205738 | DQ205851 |
| | *Arctocephalus forsteri* New Zealand fur seal | DQ205888 | DQ205777 | DQ205811 | | |
| | *Arctocephalus gazella* Antarctic fur seal | DQ205891 | DQ205780 | DQ205814 | | |
| | *Otaria byronia* South American sea lion | DQ205889 | DQ205778 | DQ205812 | DQ205739 | DQ205852 |
| | *Eumetopias jubatus* Steller's sea lion | DQ205890 | DQ205779 | DQ205813 | DQ205740 | DQ205853 |
| Odobenidae | *Odobenus rosmarus* Walrus | DQ205892 | DQ205781 | DQ205815 | DQ205741 | DQ205854 |
| Phocidae | *Cystophora cristata* Hooded seal | DQ205893 | DQ205782 | DQ205816 | DQ205742 | DQ205855 |
| | *Erignathus barbatus* Bearded seal | DQ205894 | DQ205783 | DQ205817 | DQ205743 | DQ205856 |
| | *Pagophilus groenlandicus* Harp seal | DQ205901 | DQ205790 | DQ205825 | DQ205750 | DQ205863 |
| | *Halichoerus grypus* Grey seal | DQ205902 | DQ205791 | DQ205826 | DQ205751 | DQ205864 |
| | *Pusa hispida* Ringed seal | DQ205899 | DQ205788 | DQ205823 | | DQ205862 |
| | *Phoca vitulina* Harbour seal | | | | DQ205752 | DQ205865 |
| | *Phoca largha* Spotted seal | DQ205904 | DQ205793 | DQ205827 | DQ205754 | DQ205867 |
| | *Pusa caspica* Caspian seal | DQ205905 | | DQ205828 | DQ205755 | |
| | *Ommatophoca rossii* Ross seal | DQ205900 | DQ205789 | DQ205824 | DQ205749 | |
| | *Hydrurga leptonyx* Leopard seal | DQ205895 | | DQ205818 | DQ205744 | DQ205857 |
| | *Lobodon carcinophagus* Crabeater seal | DQ205896 | DQ205784 | DQ205819 | DQ205745 | DQ205858 |
| | *Leptonychotes weddellii* Weddell seal | DQ205903 | DQ205792 | | DQ205753 | DQ205866 |

**Table 2-1.** *Continued.*

| Family | Species | *IRBP* Exon I | *FES* | *GHR* | *CHRNA1* | *RHO* |
|---|---|---|---|---|---|---|
| Phocidae continued | *Mirounga angustirostris* Northern elephant seal | DQ205897 | DQ205785 | DQ205820 | DQ205746 | DQ205859 |
| | *Monachus schauinslandi* Hawaiian monk seal | DQ205898 | DQ205786 | DQ205821 | DQ205747 | DQ205860 |
| | *Monachus monachus* Mediterranean monk seal | | DQ205787 | DQ205822 | DQ205748 | DQ205861 |
| Canidae | *Canis familiaris* Dog | DQ205906 | DQ205794 | DQ205835 | DQ205756 | DQ205868 |
| | *Canis lupus* Wolf | DQ205907 | DQ205795 | DQ205836 | DQ205757 | DQ205869 |
| | *Alopex lagopus* Arctic fox | DQ205908 | DQ205796 | DQ205837 | DQ205758 | DQ205870 |
| | *Vulpes velox* Swift fox | DQ205909 | | DQ205838 | | |
| Felidae | *Felis catus* Domestic cat | Z11811* | | DQ205829 | DQ205759 | DQ205871 |
| | *Lynx canadensis* Canada lynx | DQ205910 | DQ205797 | DQ205830 | DQ205760 | DQ205872 |
| | *Puma concolor* Cougar | DQ205911 | | DQ205831 | DQ205761 | DQ205873 |
| | *Panthera tigris* Tiger | DQ205912 | | DQ205832 | DQ205762 | DQ205874 |
| | *Lynx rufus* Bobcat | DQ205913 | | DQ205833 | DQ205763 | DQ205875 |
| | *Panthera leo* Lion | DQ205914 | | DQ205834 | DQ205764 | DQ205876 |

**Table 2-2.** General sequencing results by data partition.

| Gene partition | Number of taxa | Length (bp) | Number of informative sites excluding informative gaps | Number of informative gaps |
|---|---|---|---|---|
| *FES* | 57 | 454 | 171 | 10 |
| *GHR* | 65 | 652 | 220 | 13 |
| *RHO* | 62 | 280 | 83 | 3 |
| *CHRNA1* | 63 | 394 | 156 | 12 |
| *IRBP* | 71 | 1194 | 314 | 1 |
| All STS | 67 | 1780 | 630 | 38 |
| All data | 84 | 2974 | 949 | 39 |

**Table 2-3.** Maximum parsimony (MP) and maximum likelihood (ML) results by data partition.

| Gene partition | MP Tree Length | Number of MP Trees | Consistency Index | Retention Index | ML Score (-lnL) | Model |
|---|---|---|---|---|---|---|
| *FES* | 489 | 59328 | 0.675 | 0.903 | 3105.76606 | HKY+G |
| *GHR* | 578 | 92600 | 0.73 | 0.917 | 4004.29335 | K80+I+G |
| *RHO* | 258 | 1583 | 0.593 | 0.863 | 1666.39816 | K80+I+G |
| *CHRNA1* | 446 | 7184 | 0.704 | 0.907 | 2798.98509 | K80+G |
| *IRBP* | 905 | 48 | 0.627 | 0.895 | 6567.5753 | HKY+I+G |
| All STS | 1792 | 15676 | 0.68 | 0.9 | 11958.2738 | K81 G |
| All data | 2708 | 92600 | 0.66 | 0.898 | 18759.1817 | TVMef+I+G* |

*Transversional model equal frequencies (D. Posada, ModelTest).

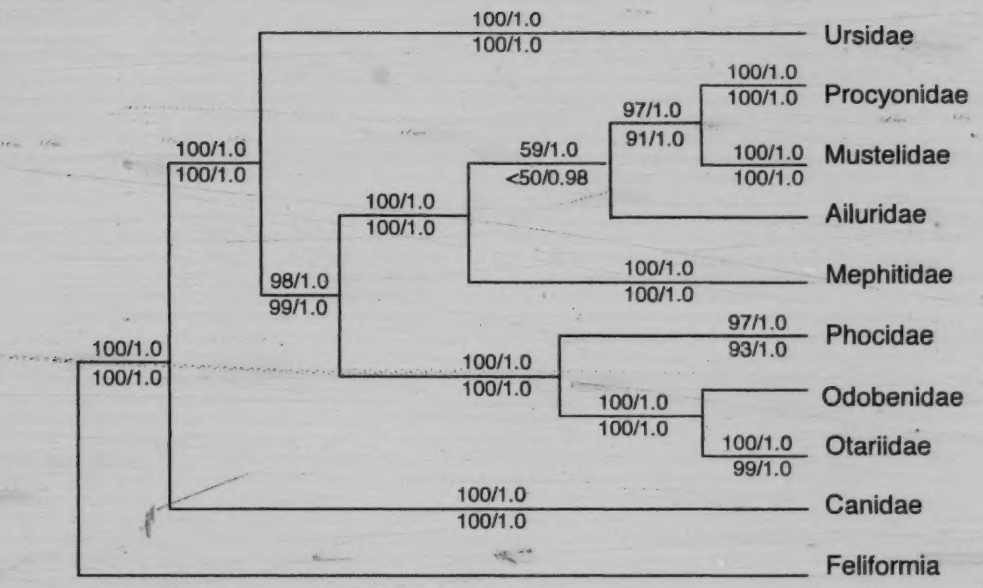


**Figure 2-1.** Family level phylogeny of the Caniformia. Maximum parsimony bootstrap / Bayesian posterior probability values are listed above the branch for the data set containing all genes and below the branch for the data set containing only the STS introns. All supermatrix and supertree methods recovered the same family-level topology, except the ML-input supertree, which could not resolve the position of the red panda (Ailuridae) within the musteloids.

**Figure 2-2. (on following page)** Maximum likelihood tree recovered from the combined data set containing five nuclear genes (-*ln*L=18759.18170). Branch lengths are optimized and correspond to the number of substitutions/site indicated by the scale bar. The TVMef + I + Γ substitution model (transversional model, equal frequencies) was implemented, as selected by ModelTest 3.06. Base frequencies were therefore set to be equal, the proportion of invariant sites = 0.229823, gamma shape = 0.932748, and substitution rates were: A-C= 1.07832, A-G=C-T= 5.24403, A-T= 0.58459, C-G= 0.97498, G-T=1.0. Bars along the right-hand side of the tree indicate family and higher-level taxonomic groups.

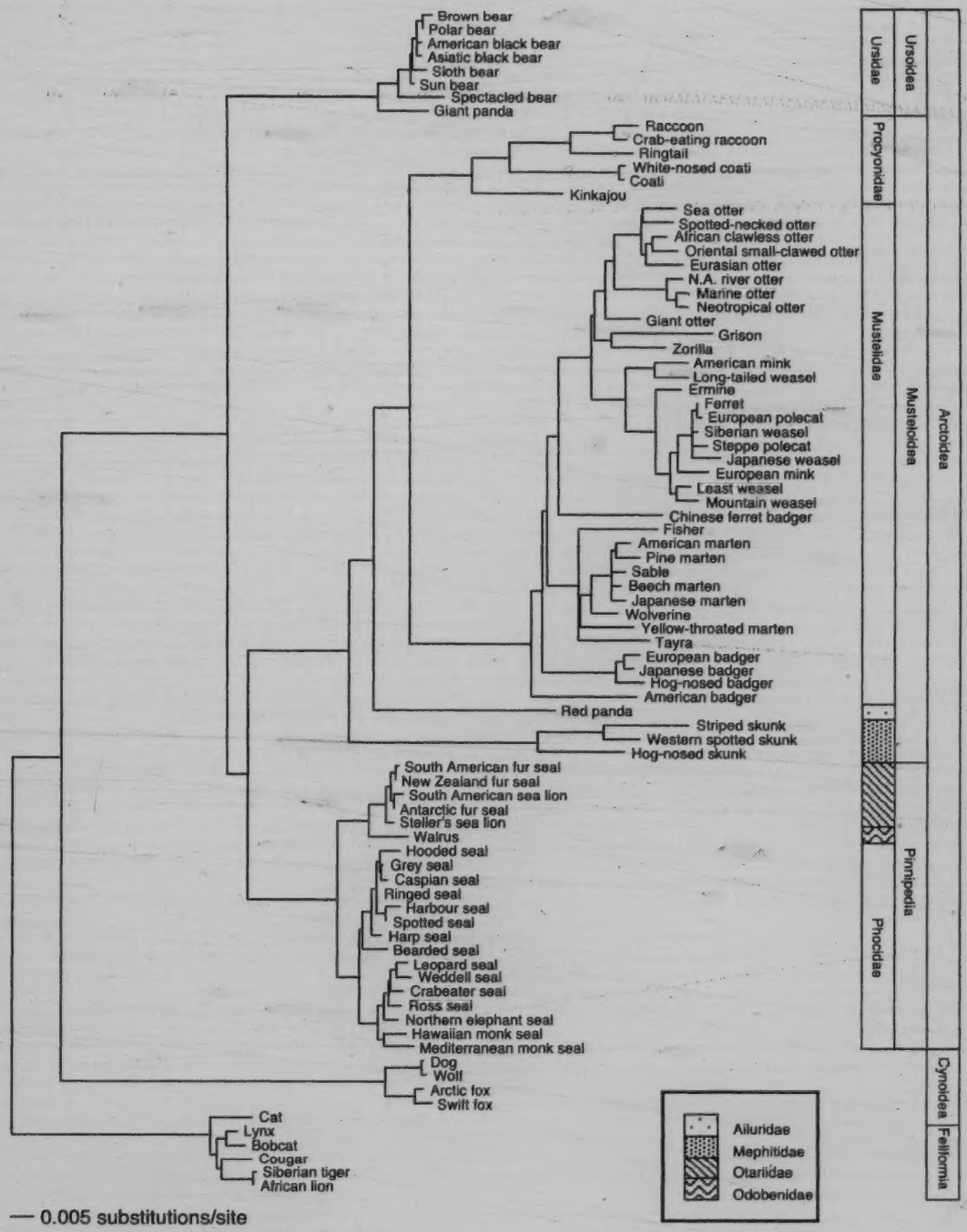


**Figure 2-2.** Figure legend on previous page.

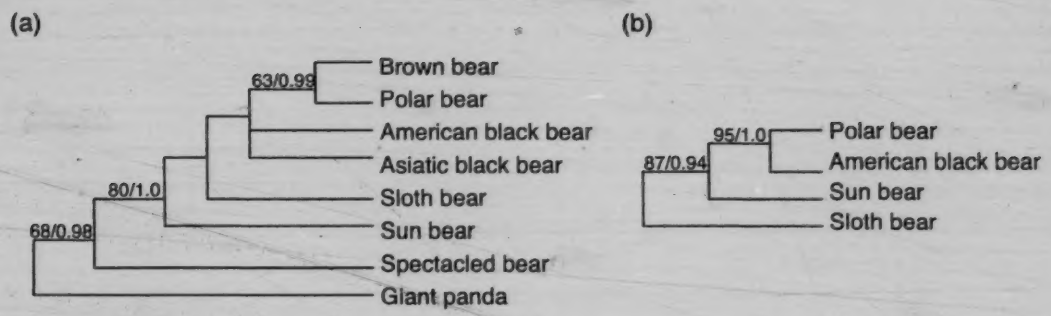


**Figure 2-3 a,b.** Maximum likelihood cladogram for Ursidae, recovered (a) from all genes and (b) from STS introns only. Support shown is MP BP/BPP for each data set. Nodes that do not have support values indicated were polytomous in both the MP BP and Bayesian topologies. The MP BP tree for all genes additionally resolved the American black bear as closer to the brown and polar bears (59% MP BP). The Bayes-input and ML-input supertrees were identical to the Bayesian topology found using all genes (i.e. resolving only three nodes), while the MP-input supertree additionally resolved the sun bear in the same position as in (a).

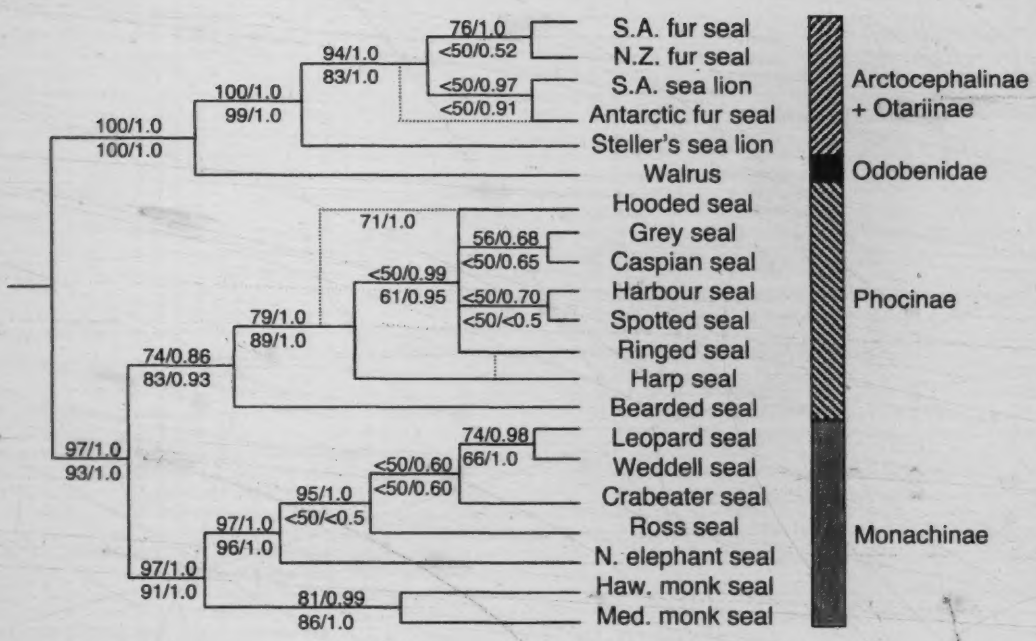


**Figure 2-4.** Phylogeny of the Pinnipedia. The maximum likelihood cladogram is shown with the MP BP/BPP support from the data set with all genes (above the branch) and from STS only (below the branch). Values of <50% MP BP or <0.5 BPP indicate polytomies in the respective MP BP or Bayesian trees. The MP-input and Bayes-input supertrees recovered consistent topologies but were less resolved. The dotted line indicates the topology (with support) recovered using the STS-only data set, placing the hooded seal as more basal than the harp seal. This pattern was also observed in the ML-input supertree, which also placed the ringed and harp seals as sister and the Antarctic fur seal as more basal than the South American sea lion (indicated by dotted branches). Subfamily designations are indicated on the right.

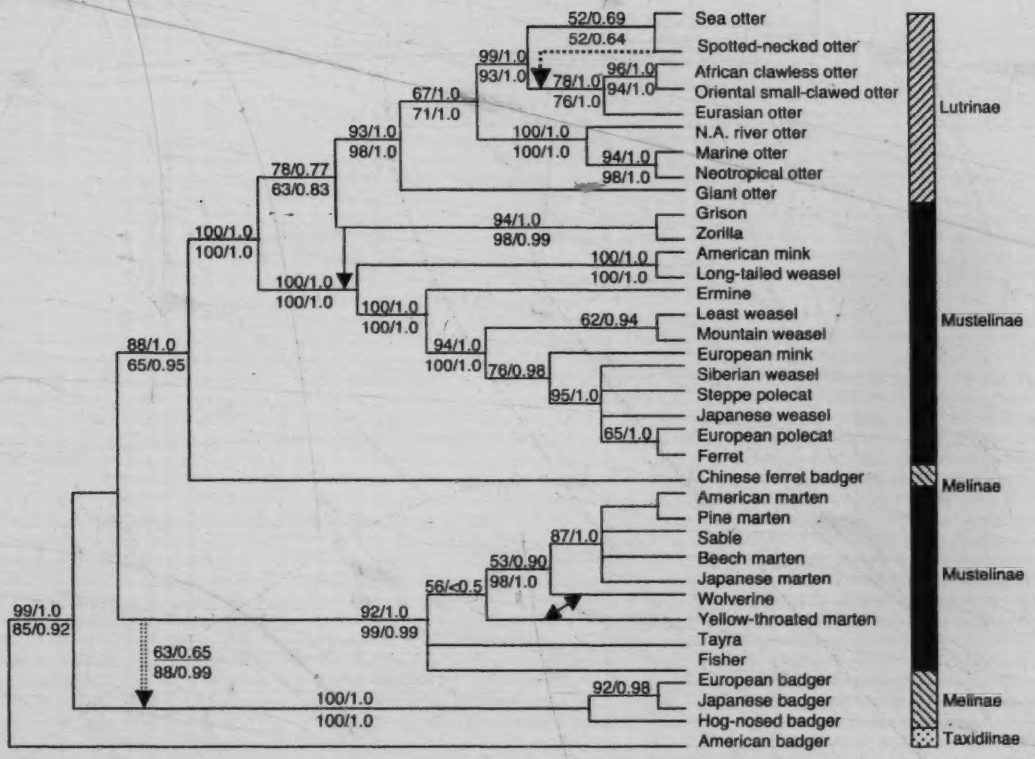


**Figure 2-5.** Phylogeny of the Mustelidae adapted from the maximum likelihood topology (Fig 2-1). MP BP / BPP values are listed above the branch for the all-data set and below the branch for the STS-only data set. Nodes that do not have support values indicated were polytomous in both the MP BP and Bayesian topologies. The above topology was recovered by all supertree and supermatrix methodologies with the following exceptions. The double-dotted arrow indicates (with associated support) the placement of the *Martes-Eira-Gulo* clade as sister to Melinae, as recovered by Bayesian, MP BP, and the MP-input supertree (ML-input and Bayes-input supertrees were unresolved). The solid arrow indicates the sister relationship between the grison+zorilla clade and the *Mustela* clade, while the double-headed arrow indicates the reciprocal placement of the wolverine and the yellow-throated marten; both as recovered by the ML-input supertree. The dotted arrow indicates the sequential branching of the sea otter and spotted-necked otters (vs. sister) recovered in the MP-input supertree. Bars along the right-hand side indicate subfamily designations.

**References**


Alfaro, M., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255-266.

Anderson, E., 1989. The phylogeny of mustelids and the systematics of ferrets. In: Seal, U.S., Thorne, E.T., Bogan, M.A., Anderson, S.H., (Eds.), Conservation biology and the black-footed ferret. Yale University Press, New Haven, Connecticut, pp. 10-20.

Baum, B. R., 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41, 3-10.

Baum, B. R., Ragan, M. A., 2004. The MRP method. In: Bininda-Emonds, O.R.P., ed. Phylogenetic supertrees: combining information to reveal the tree of life. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 17-34.

Bininda-Emonds, O. R. P., 2004. The evolution of supertrees. Trends Ecol. Evol. 19, 315-322.

Bininda-Emonds, O. R. P., Gittleman, J. L., Purvis, A., 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. 74, 143-175.

Bininda-Emonds, O. R. P., Sanderson, M. J., 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. Syst. Biol. 50, 565-579.

Brunner, S., 2000. Cranial morphometrics of fur seals and sea lions (Family: Otariidae) - systematics, geographic variation and growth. Veterinary Anatomy and Pathology. University of Sydney, NSW 2006, Sydney, 254 pp.

Bryant, H. N., Russell Fls, A. P., Fitch, W. D., 1993. Phylogenetic relationships within the extant Mustelidae (Carnivora): appraisal of the cladistic status of the Simpsonian subfamilies. Zool.J.Linn.Soc. 108, 301-334.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., Thompson, J. D., 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497-3500.

Davis, C. S., Delisle, I., Stirling, I., Siniff, D. B., Strobeck, C., 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. Mol. Phylogenet. Evol. 33, 363-377.

Decker, D. M., Wozencraft, W. C., 1991. Phylogenetic analysis of recent procyonid genera. J. Mammal. 72, 42-55.

Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Mol. Phylogenet. Evol. 37, 192-201.

Domingo-Roura, X., Lopez-Giraldez, F., Saeki, M., Marmi, J., 2005. Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores. Genet.Res.Cambridge 85, 223-233.

Dragoo, J. W., Honeycutt, R. L., 1997. Systematics of Mustelid-like carnivores. J. Mammal. 78, 426-443.

Farris, J. S., Källersjö, M., Kluge, A. G., Bult, C., 1995. Constructing a significance test for incongruence. Syst. Biol. 44, 570-572.

Flynn, J. J., Finarelli, J. A., Zehr, S., Hsu, J., Nedbal, M. A., 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst. Biol. 54, 317-337.

Flynn, J. J., Nedbal, M. A., 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs incompatibility among multiple data sets. Mol. Phylogenet. Evol. 9, 414-426.

Flynn, J. J., Nedbal, M. A., Dragoo, J. W., Honeycutt, R. L., 2000. Whence the red panda? Mol. Phylogenet. Evol. 17, 190-199.

Flynn, J. J., Neff, N. A., Tedford, R. H., 1988. Phylogeny of the Carnivora. In: Benton, M.J., ed. The Phylogeny and Classification of the Tetrapods. Clarendon Press, Oxford, pp. 73-116.

Flynn, J. J., Wesley-Hunt, G. D., 2005. Carnivora. In: Rose, K.D., Archibald, J.D., (Eds.), The rise of placental mammals. The John Hopkins University Press, Baltimore, Maryland, pp. 175-198.

Fyler, C. A., Reeder, T. W., Berta, A., Antonelis, G., Aguilar, A., Androukaki, E., 2005. Historical biogeography and phylogeny of monachine seals (Pinnipedia: Phocidae) based on mitochondrial and nuclear DNA data. J. Biogeography 32, 1267-1279.

Goldman, N., Anderson, J. P., Rodrigo, A. G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652-670.

Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47, 9-17.

Huelsenbeck, J. P., 1995. The performance of phylogenetic methods in simulation. Syst. Biol. 44, 17-48.

Huelsenbeck, J. P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Koepfli, K.-P., Wayne, R. K., 1998. Phylogenetic relationships of otters (Carnivora: Mustelidae) based on mitochondrial cytochrome *b* sequences. J. Zool. Lond. 246, 401-416.

Koepfli, K.-P., Wayne, R. K., 2003. Type I STS markers are more informative than cytochrome *b* in phylogenetic reconstruction of the Mustelidae (Mammalia: Carnivora). Syst. Biol. 52, 571-593.

Ledje, C., Arnason, U., 1996a. Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora with particular emphasis on the Caniformia. J. Mol. Evol. 42, 135-144.

Ledje, C., Arnason, U., 1996b. Phylogenetic relationships within caniform carnivores based on analyses of the mitochondrial 12S rRNA gene. J. Mol. Evol. 43, 641-649.

Nowak, R. M., 1991. Walker's Mammals of the World. Fifth Edition, 5 edn The John Hopkins University Press, Baltimore.

Posada, D., Crandall, K. A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50, 580-601.

Ragan, M. A., 1992. Phylogenetic inference based on matrix representation of trees. Mol.Phylogenet.Evol. 1, 53-58.

Rambaut, A., Drummond, A., 2003. Tracer, version v1.0.1.

Rannala, B., Huelsenbeck, J. P., Yang, Z., Nielson, R., 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47, 702-710.

Ronquist, F., Huelsenbeck, J. P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Sato, J. J., Hosoda, T., Wolsan, M., Suzuki, H., 2004. Molecular phylogeny of Arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of the ferret-badgers and skunks. Zool. Sci. 21, 111-118.

Sato, J. J., Hosoda, T., Wolsan, M., Tsuchiya, K., Yamamoto, M., Suzuki, H., 2003. Phylogenetic relationships and divergence times among Mustelids (Mammalia: Carnivora) based on nucleotide sequences of the Nuclear Interphotoreceptor Retinoid Binding Protein and mitochondrial cytochrome *b* genes. Zool. Sci. 20, 243-264.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114-1116.

Simpson, G. G., 1945. The principles of classification and a classification of mammals. Bull.Am.Mus.Nat.Hist. 85, 1-350.

Stanhope, M. J., Czelusniak, J., Si, J.-S., NIckerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Binding Protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. 1, 148-160.

Steel, M. A., Dress, A., Böcker, S., 2000. Simple but fundamental limitations on supertree and consensus tree methods. Syst. Biol. 49, 363-368.

Sullivan, J., Abdo, Z., Joyce, P., Swofford, D. L., 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. Mol. Biol. Evol. 22, 1386-1392.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). (Sunderland, Massachusetts, Sinauer Associates).

Talbot, S. L., Shields, G. F., 1996. A phylogeny of the bears (Ursidae) inferred from complete sequences of three mitochondrial genes. Mol. Phylogenet. Evol. 5, 567-575.

Templeton, A. R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221-244.

Thorley, J. L., Page, R. D. M., 2000. RadCon: Phylogenetic tree comparison and consensus. Bioinformatics 16, 486-487.

van Zyll de Jong, C. G., 1991. A brief review of the systematics and a classification of the Lutrinae. Habitat 6, 79-83.

Vrana, P. B., Milinkovitch, M. C., Powell, J. R., Wheeler, W. C., 1994. Higher level relationships of the arctoid Carnivora based on sequence data and "total evidence". Mol. Phylogenet. Evol. 3, 47-58.

Waits, L., Sullivan, J., O'Brien, S. J., Ward, R. H., 1999. Rapid radiation events in the family Ursidae indicated by likelihood phylogenetic estimation from multiple fragments of mtDNA. Mol. Phylogenet. Evol. 13, 82-92.

Wiens, J. J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol. 52, 528-538.

Wiens, J. J., 2006. Missing data and the design of phylogenetic analyses. J.Biomed.Inf. 39, 34-42.

Wolsan, M., 1993. Phylogeny and classification of early European Mustelida (Mammalia, Carnivora). Acta Theriol. 38, 345-384.

Wozencraft, W. C., 1993. Order Carnivora. In: Wilson, D.E., Reeder, D.A., (Eds.), Mammal species of the world: a taxonomic and geographic reference. Smithsonian Institution Press, Washington, DC, pp. 279-348.

Wynen, L. P., Goldsworthy, S. D., Insley, S. J., Adams, M., Bickham, J. W., Francis, J., Gallo, J. P., Hoetzel, A. R., Majluf, P., White, R. W. G., Slade, R. W., 2001. Phylogenetic relationships within the eared seals (Otariidae: Carnivora): Implications for the historical biogeography of the family. Mol. Phylogenet. Evol. 21, 270-284.

Wyss, A. R., Flynn, J. J., 1993. A phylogenetic analysis and definition of the Carnivora. In: Szalay, F., Novacek, M., McKenna, M., (Eds.), Mammal Phylogeny: Placentals. Springer-Verlag, New York, pp. 32-52.

Yu, L., Li, Q.-w., Ryder, O. A., Zhang, Y.-p., 2004a. Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. Mol. Phylogenet. Evol. 32, 480-494.

Yu, L., Li, Q.-w., Ryder, O. A., Zhang, Y. P., 2004b. Phylogenetic relationships within mammalian order Carnivora indicated by sequences of two nuclear DNA genes. Mol. Phylogenet. Evol. 33, 694-705.

## Chapter 3

## Locus congruence and conflict and the effects of taxon sampling on elucidating the phylogenetic position of the red panda

### Introduction

The evolutionary affinity of the red panda (*Ailurus fulgens*) has remained one of the mysteries of arctoid carnivore phylogenetics. *Ailurus* has been considered a close relative of the giant panda, *Ailuropoda melanoleuca*, alone or with the bear family, Ursidae; a member of the raccoon family, Procyonidae; or simply regarded as its own unique lineage within Arctoidea (for review, see Roberts and Gittleman, 1984; Flynn et al., 2000). *Ailurus fulgens* is now considered to be the sole extant species of the family Ailuridae, belonging to the Arctoidea (Carnivora, Mammalia), along with the bear family (Ursidae), the seal, sea lion, fur seal, and walrus group (Pinnipedia), the weasel, badger, and otter family (Mustelidae), raccoon family (Procyonidae), and skunk family (Mephitidae). Ailuridae is associated with the Mustelidae, Procyonidae, and Mephitidae within the superfamily Musteloidea (Flynn et al., 2000; Delisle and Strobeck, 2005; Flynn et al., 2005; Fulton and Strobeck, 2006; Sato et al., 2006; Arnason et al., 2007; Yonezawa et al., 2007; Finarelli, 2008). Historically, the skunk family was included as a subfamily in Mustelidae; its elevation to familial status as Mephitidae (Ledje and Arnason, 1996; Dragoo and Honeycutt, 1997) shed considerable light on the phylogenetic affinity of the red panda (Flynn et al., 2000). The Mustelidae *sensu stricto* (*sine* skunks) are consistently resolved as the sister group to the Procyonidae, with Ailuridae and Mephitidae placed more basally.

All three possible relationships between Ailuridae, Mephitidae, and Procyonidae + Mustelidae have been recently proposed using molecular data (Figure 3-1). A sister relationship between Ailuridae and Mephitidae was first proposed based on parsimony and likelihood analyses of 1 nuclear and 3 mitochondrial (mt) markers (Flynn et al., 2000) and has subsequently been recovered using near-complete or complete mt genomes (Delisle and Strobeck, 2005; Arnason et al., 2007) and parsimony analysis of 3 nuclear + 3 mt loci (Fulton and Strobeck, 2007). Use of likelihood or Bayesian methodology of two different sets of 6 combined nuclear and mitochondrial genes recovered the red panda as the first branching lineage (Flynn et al., 2005; Fulton and Strobeck, 2007), as did differently modeling mt genome evolution either by use of amino acid sequence or excluding or recoding third codon positions (Arnason et al., 2007). Likelihood analysis of microsatellite flanking region sequence also weakly supported Ailuridae as

basal (Domingo-Roura et al., 2005). The placement of Mephitidae as the basal lineage was proposed based on analyses of 3-5 nuclear genes (Fulton and Strobeck, 2006; Sato et al., 2006). Extensive maximum likelihood analyses of mt genomes plus 8 nuclear genes also recovered this Mephitidae-basal relationship, although there was nearly equal support for an Ailuridae-Mephitidae sister grouping (Yonezawa et al., 2007). While Bayesian posterior clade probabilities are generally high in studies recovering either of the latter two hypotheses, bootstrap support is generally low (but see Sato et al., 2006) and a consensus has not yet been reached.

The difficulty in resolving the phylogenetic position of the red panda and the apparent conflict between studies is likely resultant from the rapid radiation between the three musteloid lineages, Ailuridae, Mephitidae, and Mustelidae+Procyonidae. In such a rapid radiation, strong incongruence between loci may result from deep coalescent events (incomplete lineage sorting) or artifactual conflict due to violations of model assumptions such as variation in rate or base composition between lineages. Often, these problems are confounded by insufficient time for enough mutations to arise on short branches, providing very little phylogenetic information regarding branching order. Weak phylogenetic signal along short branches joining long branches together can often result in long-branch artifact problems. Here, I present 7 new nuclear loci in combination with 7 other nuclear genes and near-complete mitochondrial genomes previously examined by myself or others, increasing the number of independent loci of varying evolutionary rate and potentially obtain more phylogenetic signal across these loci to resolve this rapid radiation. With an increase in the number of loci comes increased potential for incongruence. Factors that can lead to artifactual resolution are investigated, as are the effects of applying methods that may compensate for them, such as exclusion of partitions that are in conflict, those that exhibit biases, and the effects of taxon sampling and outgroup selection. The identification of possible causes of conflict between loci may yield insight into the previous discrepancy between studies and provide a comprehensive view of the evolution of the red panda.

## Materials and methods

### *Samples and sequencing*

Thirty-three taxa, comprised of 2 canids, 2 felids, 5 pinnipeds, 6 ursids (including the giant panda), 4 procyonids, 3 mephitids, 10 mustelids, and *Ailurus*, were included in this study (Table 3-1). Sequences for 14 nuclear loci and 12 mitochondrial protein-coding genes were

obtained and PCR primer sequences and references are listed in Table 3-2. Sequences for *IRBP, FES, GHR, CHRNA1*, and *RHO* are from Chapter 2. Sequences for these five genes for *P. vitulina*, *H. leptonyx*, and *M. angustirostris* are from Chapter 5. Sequences for the remaining nuclear genes were primarily obtained by this study and are marked as newly generated in Table 3-3. Information for any sequences that were downloaded from GenBank is also included. Amplification and sequencing strategies are described in detail in Chapter 5. 11 new near-complete mitochondrial genomes (12 protein-coding genes) are presented here for the sloth bear, white-nosed coati, ringtail, kinkajou, western spotted skunk, hog-nosed skunk, sea otter, European badger, domestic ferret, ermine, and least weasel. Sequences were obtained following the strategy of Delisle and Strobeck (2002), with new primer design when necessary. For several species, cytochrome *b* was not sequenced and sequences were obtained from GenBank. Mitochondrial gene accession numbers and references are listed in Table 3-4.

*Sequence alignment, marker congruence, and model selection*

*IRBP, GHR, CHRNA1, RHO*, and *FES* alignments from Chapter 2 were used (Fulton and Strobeck, 2006). Sequences for *PLCB4* and *FLVCR1* were aligned using MAFFT v.6.240 (Katoh et al., 2002; Katoh et al., 2005) under the default settings for FFT-NSi (fast Fourier transform, iterative refinement) and adjusted manually. All other new loci had few to no insertion-deletion (indel) events and were aligned by eye, with indels in frame where applicable.

Congruence of each nuclear locus with the other nuclear loci was tested using the incongruence length difference (ILD) test (Farris et al., 1995) implemented as the partition homogeneity test in PAUP* v.4.0b10 (Swofford, 2003).

The best-fit model of DNA evolution was selected for each nuclear locus separately and combined and for the mtDNA dataset, using the AIC criterion in MrModelTest v.2.2 (Nylander, 2004), which tests a restricted set of models from ModelTest (Posada and Crandall, 1998).

For the mtDNA dataset, the amino acid sequence was inferred using Mesquite OSX v.1.12 (Maddison and Maddison, 2006), with stop codons excluded, but overlapping regions between genes included in each gene, due to different reading frames.

*Base composition bias and saturation*

Base composition homogeneity across taxa was assessed in PAUP* for each nuclear

locus separately and combined, and for each mt gene separately and combined. The concatenated mt dataset was partitioned by codon position and each codon position was also assessed for base composition bias.

To visualize the level of saturation in the mtDNA dataset, pair-wise comparisons between all taxa were made and uncorrected p-distance was plotted against the transition-transversion ratio (Ti:Tv). This was performed for both the entire dataset and a dataset excluding 3rd positions. For comparison, this was also performed for the primarily non-coding sequence from *FES* and the coding sequence of *BRCA1*. In some cases within Ursidae, no transversions were observed, giving an undefined Ti:Tv ratio. As the Ursidae only represent outgroups in this study, those points were omitted from the plot.

*Relative rate tests*

Relative rate tests were implemented in GRate (Muller, 2002). Based on the work of Sarich and Wilson (1967), GRate implements a likelihood-based approach, allowing user-selected models of evolution and user-defined lineages for comparison. Standard error is obtained using bootstrapping and the significance of the result is evaluated using a two-tailed *z*-test, including 95% confidence intervals. For each nuclear gene, the best-fit model selected from MrModelTest was employed; for combined nuclear analyses and for mtDNA (with 3rd positions excluded), GTR+I+$\Gamma_4$ was used. 100 bootstrap replicates were used to obtain standard error estimates. Felidae was set as the outgroup, with the remaining taxa divided into primarily familial lineages: Canidae, Pinnipedia, Ursidae, Mustelidae, Procyonidae, Mephitidae, and Ailuridae.

*Tree search analyses*

Several analyses were implemented in the same manner for all datasets including each nuclear gene, all nuclear genes combined, all nuclear genes excluding *RAG1* and with canids excluded from *IRBP*, the nuclear dataset excluding those genes that illustrated rate differences between several lineages within Musteloidea (*ADORA3, APOB, BRCA1*), the nuclear dataset excluding the most strongly supported 'opposing' genes (*FES* and *PLCB4*), mtDNA, mtDNA excluding 3rd positions, and mt inferred amino acid sequence. Maximum parsimony (MP) and minimum evolution (ME) using LogDet distance (Lockhart et al., 1994) searches were implemented in PAUP* (Swofford, 2003). For MP, 10 replicates of random addition starting

trees were used; for ME, neighbour-joining starting trees were used. 100 bootstrap pseudoreplicates (with 10 random addition starting trees per pseudoreplicate) were used for both MP and ME. Maximum likelihood (ML) analyses were performed in RAxML v.7.0.0 (Stamatakis, 2006a, b). For combined nuclear analyses, each locus partition was allowed its own parameters within the GTR+Γ model for both tree searching and rapid bootstrapping (100 replicates, GTR+CAT approximation model). MtDNA was analyzed as a single partition, by codon position (3 partitions), by gene (12 partitions), and by gene and codon position (36 partitions). Mt amino acid sequence was analyzed with a separate MTMAM model (mammalian mitochondrial amino acid replacement matrix) applied to each gene and empirical base frequencies (F) used. The mt amino acid dataset was also analyzed in PhyloBayes v.2.3 (Lartillot et al., 2007), which is a Bayesian MCMC based method that implements a mixture model (CAT) for the substitution processes at each site in the alignment (Lartillot and Phillipe, 2004). Two independent runs were performed using the CAT-POISSON model and convergence was determined using the PhyloBayes *bpcomp* program. All trees were viewed in either TreeView X (Page, 1996) or FigTree v.1.1.2 (Rambaut, 2008).

SplitsTree4 v.4.8 (Huson and Bryant, 2006) was used to construct a NeighbourNet diagram using logdet distance from the concatenated nuclear dataset.

*Effects of Mephitidae taxon sampling and outgroup selection*

To determine if the level of taxon sampling within Mephitidae had an effect on the recovered topology, the same analysis was run including 1, 2, or 3 mephitid genera. For this test, mtDNA was analyzed under MP, ME, and ML as a single partition for all combinations of mephitid taxa and 100 replicates of bootstrapping performed.

ME, MP, and ML analyses were performed for the nuclear dataset (partitioned by gene for ML analyses) to determine the effects of outgroup selection. All musteloids were included with either pinnipeds, ursids, canids, or felids as the only outgroup, and pinnipeds+ursids, or pinnipeds+ursids+canids.

*Hypothesis testing*

Consel v.0.1i (Shimodaira and Hasegawa, 2001) was used to implement the approximately unbiased (AU) test (Shimodaira, 2002) and Shimodaira-Hasegawa (SH) test

(Shimodaira and Hasegawa, 1999). Three topologies representing the three alternate placements of the red panda (Figure 1) were tested. The ferret, European badger, and sun bear were excluded from these trees and tests to ensure that any significance (or lack thereof) was due to alternate placement of the red panda and not the other three taxa that occasionally are recovered in different positions. The mtDNA dataset excluding 3rd positions, each nuclear gene, and all nuclear genes (excluding incongruent partitions) were tested. The best-fit model of evolution for each locus/dataset was used to calculate the site likelihoods.

**Results**

*Sequencing results*

As in Chapter 2, sequences for the cat and striped skunk are missing for *FES*, and the giant panda for *RHO*. Due to lack of a DNA sample, giant panda sequences for *RAG2, PNOC, ADORA3, APOB, FLVCR1*, and *PLCB4* could not be obtained. *RAG2* amplification could not be obtained for canids, *RAG1* for the sun bear, or *BRCA1* for the kinkajou.

Some insertion-deletion (indel) events were observed in the mt genes. In cytochrome b, the Steller's sea lion has a 2 amino acid (aa) insertion (ACGGCT: Thr-Ala) near the 3' end, as does the wolverine (ACACTT: Thr-Leu). Immediately before the cytochrome b stop codon, the giant panda has a 3aa insertion (TGACTTCCA: Trp-Leu-Pro). The sea and river otters share a 1aa deletion immediately before the *ATP8* stop codon. Also in *ATP8*, the kinkajou has a 1aa (TTA: Leu) insertion immediately preceding the stop codon. *COI* is 2 aa longer in the kinkajou and *CYTB* is 2 aa longer in the white-nosed coati, as described in Chapter 4 (Fulton and Strobeck, 2007). Both the Northern elephant seal and walrus have a single aa insertion (CAA: Gln) at the end of *COII*.

As described in Chapter 2 (Fulton and Strobeck, 2006), a 204 base pair (bp) insertion was removed from *GHR* in both canids and a 204 bp insertion in *RHO* was removed from the red panda alignment. A 221 bp insertion in *FLVCR1* in both felids resembling a SINE insertion (see Chapter 5) was removed. The kinkajou was heterozygous for an indel in *FLVCR1*; the sequence could not be determined and was annotated as missing in the matrix. The mink was heterozygous for a 3 bp indel in *PLCB4*. The sequence of the longer allele was determined and this allele is included to represent the species sequence.

No indels conferring phylogenetic information regarding the position of the red panda within Musteloidea were found. However, a 25 bp deletion in *GHR* and a 12 bp deletion in *PLCB4* were shared by all musteloid taxa, including the red panda. Many other indels were found to support other widely recognized clades.

*Locus congruence*

*IRBP, FES, CHRNA1,* and *BRCA1* were rejected as congruent with the other nuclear loci by the ILD test at $p$=0.05. *RAG1* was rejected at $p$=0.01. With Bonferroni correction for multiple tests (14 tests), the significance level of 0.05 would be reduced to 0.003. Given the many problems with the ILD test (Dolphin et al., 2000; Barker and Lutzoni, 2002; Darlu and Lecointre, 2002; Dowton and Austin, 2002; Quicke et al., 2007) and the suggestion that it never be used for partition combinability (Yoder et al., 2001; Barker and Lutzoni, 2002), 'rejected' loci were further investigated to determine if they illustrated any hard topological conflicts (>80% MLBP) with other loci, excluding the position of the red panda. *IRBP* strongly supported (MLBP=99%) a Canidae+Musteloidea sister grouping, which is highly incongruent with all other loci and accepted taxonomy. Once canids were excluded from the *IRBP* dataset, the partition was no longer incongruent (based on ILD testing and topological comparisons). *FES, CHRNA1*, and *BRCA1* did not show any hard topological incongruencies with other loci and were therefore retained. *RAG1* received a $p$-value of 0.01, thus, it would not be excluded after correction, but based on the low value and the topological incongruencies, *RAG1* was rejected as congruent. *RAG1* strongly supported (100% ML BP) many clades that are highly incongruent with accepted taxonomy (i.e. river otter + bears, lynx + striped skunk, coati + wolf, etc.). Thus, *RAG1* was excluded from further analyses. All analyses of the combined nuclear dataset represent *IRBP* excluding canids and completely excluding *RAG1* unless otherwise noted. Of the mt genes, only cytochrome *b* was rejected at $p$=0.05 when compared to all other genes. However, no strongly supported topological incongruencies were observed and the locus was retained.

*Base composition bias and saturation plots*

No base composition bias was indicated for any individual nuclear gene ($p$=1.0). However, the combined nuclear dataset did show base composition bias across taxa ($p$=0.00888299). The mtDNA dataset showed significant base composition bias across taxa

($p$<0.0001), but significant bias was no longer observed when the 3rd codon positions were removed ($p$=0.28161). *NADH dehydrogenase 4 (ND4)* and *NADH dehydrogenase 5 (ND5)* showed significant bias ($p$=0.00213, $p$=0.000248). This appears due to the ND4 2nd codon positions and ND5 3rd codon positions. If ND4 and ND5 are excluded, the remaining combined genes still exhibit bias across taxa ($p$<0.0001). Therefore, to reduce base composition bias across taxa, the 3rd codon positions were excluded from all mt genes.

Both mtDNA saturation plots (Figure 3-2 a, b) show a clear curvilinear trend, indicating saturation has occurred for the entire dataset for any p-distance > ~0.185 and for the 1st and 2nd positions only for any $p$-distance > ~0.09. All of the pair-wise comparisons including the red panda, whether to other musteloids or to more distantly related carnivores, fall in the range of values that appear to be saturated. For comparison, plots for two randomly-chosen nuclear genes, *FES* and *BRCA1*, were performed (Figure 3-2 c, d). Both showed a larger range of Ti:Tv values at low $p$-distances compared to higher distances, indicating that while transitions may occur slightly more frequently than transversions between closely related taxa. No other strong trend was observed. The Ti:Tv ratio was generally between 1.5 and 3.5 for *BRCA1* and 1 and 5 for *FES*.

*Rate variation across lineages*

Only *BDNF*, *PNOC*, and *RAG2* showed no significant differences between lineages using the relative rate test implemented in GRate. Other loci showed at least one incidence where two lineages had significantly different rates, though not necessarily within Musteloidea (Table 3-5). For *ADORA3*, Ailuridae was significantly different from all other families except Canidae, especially from Procyonidae ($p$<0.0001). For *APOB*, both Mephitidae and Ailuridae were different ($p$<0.05 or <0.01) from Ursidae, Pinnipedia, and Procyonidae and Procyonidae was also different from Mustelidae ($p$<0.05). Ailuridae was significantly different from Procyonidae ($p$<0.05), Mephitidae ($p$<0.01), and Mustelidae ($p$<0.001) for *BRCA1*. Procyonidae ($p$<0.05) and Ailuridae (p<0.01) were different from Mustelidae and Ailuridae from Mephitidae ($p$<0.05) for *FLVCR1*. For *PLCB4*, Ailuridae ($p$<0.01) and Mustelidae ($p$<0.05) were both significantly different from Mephitidae. Ailuridae ($p$<0.05) and Mustelidae ($p$<0.01) were different from Mephitidae for *RHO*. *FES* had one significant result within Musteloidea (Ailuridae-Mustelidae, p<0.05), but all comparisons with Canidae were highly significant ($p$<0.0001). Ailuridae and

Mustelidae were different ($p<0.05$) for *IRBP*. *GHR* and *CHRNA1* did not illustrate any significant rate differences between musteloid families.

Combining only the three loci that did not show any rate variation (*BDNF*, *PNOC*, *RAG2*) resulted in unsupported familial relationships (ML BP, results not shown). To test if rate variation between lineages was affecting the final nuclear result, the loci that exhibited the most rate variation between musteloid lineages were excluded. Loci were selected for exclusion based on meeting one of two criteria: a comparison within Musteloidea that was highly significant ($p<0.0001$) or a strong pattern that at least one musteloid lineage was significantly different from the other lineages. *ADORA3* met both criteria; *APOB* and *BRCA1* met the second. Ailuridae had a significantly different rate than all other arctoid lineages for *ADORA3* and from all other musteloid lineages for *BRCA1*; Procyonidae was significantly different from all other musteloid lineages for *APOB* (Table 3-5). These three loci were excluded from one analysis set. There was no correlation between the ML recovered topology for a locus and the level of rate variation within Musteloidea (Table 3-6).

MtDNA excluding the 3rd positions showed significant rate variation between lineages. Mephitidae was significantly different from all other families except Canidae and exhibited a slower rate in all cases (rate ratio ~0.7-0.8 in most comparisons). The ratio of rates between Mephitidae and Ailuridae was 0.5931 and was significant at $p<0.0001$. Ailuridae was also significantly different from both Mustelidae and Procyonidae at $p<0.05$ and Canidae at $p<0.01$ and was, in general, slightly faster. The only other significant result was between Canidae and Ursidae ($p<0.05$).

*Mitochondrial topologies*

Analyzing both the entire mtDNA dataset and the dataset excluding 3rd codon positions always recovered Ailuridae+Mephitidae, regardless of the analysis method employed, except logdet analysis of the mtDNA excluding 3rd positions, which recovered Ailuridae-basal (ME BP=54). When all nucleotides were included, maximum parsimony (MP) supported (MP BP=84%) Ailuridae-Mephitidae as sister to Ursidae. Both minimum evolution using logdet (ME) and maximum likelihood (ML) supported (ME BP=71, ML BP=100) Musteloidea monophyly. When the 3rd codon positions were removed, MP recovered monophyletic Musteloidea (MP BP =47). ME, MP, and ML analyses excluding mt 3rd positions (Figure 3-3) showed reduced support for Ailuridae+Mephitidae compared to the inclusion of those bases (ML: 83 to 53, ME: 58 to

topology change, MP: 61 to 55). Except ML BP support from the entire mtDNA dataset, Ailuridae+Mephitidae support was always low.

MP BP (62%), ML BP partitioned by gene (91%, Figure 3-4), and ME BP logdet (95%) analyses of the mt amino acid sequence supported Ailuridae as basal. When ME logdet analysis included 4 rate categories (versus 1), Ailuridae+Mephitidae were recovered as sister and moderately supported (ME BP=73); other familial relationships were recovered but unsupported.

*Individual nuclear gene results and tests of alternate topologies*

The best-fit model of evolution selected for each nuclear locus is listed in Table 3-6. The Musteloidea topology recovered by ML analysis and associated ML BP support are also listed in Table 3-6. Five loci, *IRBP*, *FES*, *GHR*, *BRCA1*, and *ADORA3*, recovered Mephitidae-basal. Four loci recovered Ailuridae-basal: *CHRNA1*, *RAG2*, *APOB*, and *FLVCR1*, although *FLVCR1* unconventionally grouped Mephitidae with Procyonidae (ML BP=90). The remaining four loci, *BDNF*, *PNOC*, *RHO*, and *PLCB4* recovered Mephitidae and Ailuridae as sister. Support >75% ML BP was only recovered for the Mephitidae-basal and the Mephitidae-Ailuridae sister topologies. Most ME BP and MP BP analyses of individual genes did not support any resolution of the position of the red panda. *IRBP* moderately supported (ME BP=68%, MP BP=71%) Ailuridae as sister to Mustelidae+Procyonidae (i.e. Mephitidae-basal). MP BP analysis of *BRCA1* also supported Mephitidae-basal (72%); ME recovered this topology but did not support it >50% BP. *PLCB4* MP BP (97%) and ME BP (96%) strongly supported Ailuridae and Mephitidae as sister. *FLVCR1* moderately supported Ailuridae-basal using MP (MP BP=68), but no support was obtained from ME BP and as in the ML analysis, Mephitidae and Procyonidae were sister groups.

The results of the AU and SH tests of alternate topologies are listed in Table 3-7. The probability of the best tree (the ML tree) is in bold and all probabilities <0.1 are underlined. The Ailuridae-Mephitidae sister group for *FES*, both the Ailuridae-basal and Mephitidae-basal topologies for *PLCB4*, and the Ailuridae-basal topology for *RHO* were rejected at $p<0.05$ by the AU test. No alternate topologies were rejected at $p=0.05$ using the more conservative SH test by any individual nuclear gene. Mephitidae-basal was rejected by the AU test at $p<0.1$ or $p<0.05$ by the fewest genes (*PLCB4* and *FLVCR1*). Both the Ailuridae-basal and Ailuridae+Mephitidae topologies were rejected by four genes by at least $p<0.1$ by the AU test (Table 3-7). The SH test yielded similar results, with the Mephitidae-basal topology only rejected at $p<0.1$ for the *PLCB4*

dataset, the Ailuridae-Mephitidae sister topology rejected once at $p<0.1$ using *FES*, and the Ailuridae-basal topology rejected using both of these datasets (*PLCB4* and *FES*).

The mtDNA dataset excluding 3rd codon positions could not reject either alternate topology as a possible explanation of the data (Table 3-7). However, the Ailuridae-Mephitidae sister topology was rejected at $p<0.05$ using the AU test and $p<0.1$ using the SH test for the nuclear dataset. The probability of the Ailuridae-basal topology was ~0.1 for the nuclear dataset using both the AU and SH tests.

*Combined nuclear datasets: parsimony vs. minimum evolution vs. likelihood*

ML analysis of the nuclear dataset strongly supported (ML BP=98) the placement of Ailuridae as sister to the clade Mustelidae+Procyonidae (Figure 3-5). ME with logdet distance very weakly supported this topology as well (ME BP=56), but MP recovered an Ailuridae-Mephitidae sister group (MP BP=50). Removal of the three loci exhibiting rate variation between musteloid lineages (*ADORA3*, *APOB*, *BRCA1*) had little effect. ML BP support was reduced to 88% and ME BP to 52% for the placement of Ailuridae as the second musteloid branch and the MP BP result was unchanged. Removal of the two loci that most strongly support opposing topologies (*FES* and *PLCB4*) also had little effect, where ML supported Mephitidae-basal (ML BP=90), MP BP recovered a sister relationship (MP BP=54), and ME changed to recovering the Ailuridae-Mephitidae sister relationship, but it was unsupported (ME BP=43). If *IRBP*, the only other locus illustrating >90% ML BP for the position of Ailuridae within Musteloidea, was removed, ML still recovered the Mephitidae-basal topology, but it was unsupported (ML BP=45). ME BP and MP BP analyses were also performed with the entire dataset with one locus removed (repeated for all loci). The removal of a single locus had little effect for most loci, recovering ~60% ME BP support for Mephitidae-basal and ~50% MP BP for Ailuridae+Mephitidae. Removal of *FES*, *BDNF*, or *ADORA3* resulted in ME recovery of Ailuridae+Mephitidae. Removal of *PNOC*, *PLCB4*, or *APOB* resulted in MP recovery of Mephitidae-basal. Ailuridae-basal was not recovered through analysis of the dataset minus any one gene.

*Combined nuclear and mtDNA*

When the nuclear and mt amino acid datasets were combined, there was weak ML BP support (60%) for the placement of Ailuridae as the second musteloid branch, after Mephitidae

(Mephitidae-basal topology). When the nuclear dataset was combined with mtDNA excluding 3rd codon positions, ML BP weakly supported (62%) and ME BP (logdet) moderately supported (71%) Ailuridae as the second musteloid branch, although MP BP strongly supported (90%) Ailuridae as sister to Mephitidae.

*Effects of Mephitidae taxon sampling and outgroup selection*

A reduction in MP BP support for the Ailuridae-Mephitidae clade was observed as more mephitid taxa were included (Table 3-8). As the entire mtDNA dataset was analyzed in this case, MP BP supported Ailuridae+Mephitidae as sister to Ursidae, except in the single case when only *Spilogale* was included to represent Mephitidae. ML BP support was approximately equal when either 2 or 3 mephitids were included, but increased when only 1 mephitid was included, as in MP. ME analyses showed no clear trend in support or the recovered topology based on the number of mephitids included. When only *Spilogale* was included, or when *Spilogale* and *Mephitis* were included, Ailuridae was recovered as the first musteloid lineage instead of sister to Mephitidae.

Outgroup selection did have an effect on the recovered musteloid family relationships and associated bootstrap support. Using the nuclear dataset, either Mephitidae-basal or Ailuridae+Mephitidae was recovered using ME or MP, generally with low bootstrap support, as found by the inclusion of all taxa (Table 3-9). The exception was the inclusion of only bears with musteloids, which showed much increased support for Ailuridae as the second musteloid branch. This strong support may stem from the signal that draws together skunks and bears, as observed with network analysis (Figure 3-6). ML analyses were unaffected by outgroup selection, provided any that included any arctoids other than musteloids (pinnipeds and/or ursids) were included. When neither pinnipeds nor ursids were included (only felids or canids), ML BP support was greatly reduced (Table 3-9).

*Outgroup relationships*

The nuclear, mtDNA excluding 3rd positions, mt amino acid, and combined analyses all recovered the same relationships within Pinnipedia and Mustelidae (except the relationships between the three included *Mustela* species sometimes differed). Relationships between the subset of taxa included to represent Ursidae were recovered differently depending on the dataset

used. MtDNA partitioned by both gene and codon position recovered the sun and American black bears as sister (ML BP=80), then the polar bear (ML BP=78). When the 3rd codon positions were excluded, the same topology was recovered, with slightly decreased support for the black+sun clade (ML BP=70), but increased for the sun+black+polar clade (ML BP=90). When mtDNA with 3rd positions excluded was combined with nuclear DNA, the black and polar bears were recovered, but unsupported, as sister (ML BP=18), then the sun bear is sister to the black+polar clade (ML BP=84). Mt amino acid analysis recovered the sun and polar bears as sister (ML BP=80), then the American black bear (ML BP=93); when combined with nuclear data, the topology was unchanged with similar support values. Similar to the mt with 3rd positions removed, the nuclear dataset recovered the polar and black bears as sister (ML BP=92), but very weakly supported the sun and sloth bear as sister (ML BP=52), which was not observed in any of the analyses including mt data.

## Discussion

### *Unraveling a rapid radiation*

Identifying the phylogenetic position of the red panda within Arctoidea is a problem that has persisted in carnivore systematics, as neither morphology nor DNA evidence has provided a clear picture of the evolution of this taxon. The family Ailuridae was once more speciose, with species found across both North America, Europe, and Asia (McKenna and Bell, 1997; Wang, 1997; Baskin, 1998; Sasagawa et al., 2003; Wallace and Wang, 2004; Peigne et al., 2005; Salesa et al., 2006; Sotnikova, 2008). Although studies of fossil ailurids have led to significant insights into the evolution of Ailuridae themselves and the closer relationship to musteloids than to ursids, no evidence has been found to clearly define the relationship of Ailuridae within Musteloidea (Wang, 1997; Sotnikova, 2008). Molecular studies to date have not provided much more clarity, except to strongly support the sister grouping of Mustelidae and Procyonidae, to the exclusion of Mephitidae and Ailuridae.

The lack of clarity in understanding early musteloid evolution stems from the rapid radiation of the three lineages, Ailuridae (red panda), Mephitidae (skunks), and Mustelidae (weasels, badgers, otters, etc.) plus Procyonidae (raccoons). Difficulty in resolving rapid radiations can result for several reasons. In a short time frame, few mutations occur, often resulting in little phylogenetic information. If long branches are involved and joined by these

short internal branches, several problems can occur. Along long branches, many mutations can occur, leading to homoplasy. Long branch attraction (Felsenstein, 1978) may occur based on these homoplastic characters, particularly if there is not enough phylogenetic information on the short connecting branches. Artifactual groupings can occur due to violations of model assumptions such as heterogeneity in nucleotide or amino acid composition (Galtier and Gouy, 1995; Foster, 2004; Jermiin et al., 2004), rate variation across lineages (Kuhner and Felsenstein, 1994), across sites (Kuhner and Felsenstein, 1994; Yang, 1996), or through time (e.g. Kolaczkowski and Thornton, 2004; Ruano-Rubio and Fares, 2007). Third, rapid divergences at any depth of the tree may be susceptible to incomplete lineage sorting effects (Edwards et al., 2005), leading to discordance among gene trees. Both a lack of information and conflict between genes has been shown to be problematic for short branches (Wiens et al., 2008).

*Taxon sampling and outgroup selection*

It has been proposed that increased taxon sampling within Mephitidae could yield increased resolution of familial relationships, as it may act to break up long branches (Fulton and Strobeck, 2007; Yonezawa et al., 2007). Adding taxa can help to reconcile 'long-branch' problems, provided an appropriate sampling scheme is selected (Graybeal, 1998; Poe, 2003; for review, see Heath et al., 2008). Three genera of mephitids, each represented by a single species, were included in this study (Table 3-1). Including fewer than three mephitid species generally increased both MP and ML BP support. When all three species were included, complete mtDNA analyzed as a single partition supported Ailuridae+Mephitidae with 78% ML BP support (Table 3-8). When only a single species was included, both MP and ML BP support for a sister grouping increased ~20-30% (Table 3-8). When two genera were included, the ML result did not differ from including three, except when only *Spilogale* and *Mephitis* were included. Interestingly, inclusion of only *Spilogale* did not lead to the artifactual MP grouping Ailuridae+Mephitidae with Ursidae and ME recovered the Ailuridae-basal topology as mt amino acid analysis does (Table 3-8). Including more than one taxon certainly has an effect, but the effect of including more than two is unclear. Increased taxon sampling in future studies could further clarify this relationship. More species from each included genera could be added. The genus *Mydaus* (stink badgers) was also included in the redefined Mephitidae as the basal lineage (Flynn et al., 2005). *Mydaus* has been included in previous studies (Flynn et al., 2005; Yonezawa et al., 2007), but the number of available sequences is low. Increased sequence from this basal mephitid genus may

contribute the most to breaking up the mephitid family branch to determine if taxon sampling is, in fact, a problem.

Resolving the musteloid family branching order is a case of determining the root of the musteloid tree. A very similar unrooted tree is recovered by all methods (results not shown), but the placement of the root defines the branching order. Maximum likelihood showed strong bootstrap support for Mephitidae-basal, provided any other arctoid taxa (bears or pinnipeds) were included (Table 3-9). Although the same topology was recovered, support was greatly reduced when only a more distant felid or canid outgroup was used. Both maximum parsimony and minimum evolution methods were inconsistent in their recovered topology (Table 3-9), with no clear pattern as to the effect of the relatedness of the outgroup to the ingroup, nor to the number of outgroup taxa employed.

*Identification of data biases*

Base composition bias in either nucleotide or amino acid sequence has been shown to lead to artifactual groupings whereby taxa of similar composition may be drawn together instead of groupings based on shared evolutionary history (Galtier and Gouy, 1995; Foster, 2004; Jermiin et al., 2004). When concatenated and unpartitioned, the nuclear dataset showed composition bias across taxa, as did the 3rd codon positions of the mtDNA. ML (partitioned or unpartitioned) and ME (logdet distance) of the nuclear dataset recovered the same Mephitidae-basal topology, while MP recovered a sister relationship (MP BP=54). As both ML and logdet distance are generally more robust to compositional biases than MP, the Ailuridae-Mephitidae grouping based on MP may be artifactual. This sister relationship was also recovered by mtDNA, but supported (MP BP=88) as the sister to Ursidae (Figure 3-3). As there is no precedent for this grouping from either molecular or morphological evidence, the placement of skunks with bears here is considered artifactual. This effect is only observed with MP analysis of this dataset and is remedied when the 3rd position bases are removed. All of the mt codon positions exhibit saturation for the relationships between musteloid families (Figure 3-2), thus, saturation may have also contributed to this artifactual grouping with Ursidae. This is very similar to the results of Delisle and Strobeck (2005) using a very similar mtDNA dataset, who illustrated that purine-pyrimidine (RY) coding could also correct for possible base composition bias and saturation.

On the contrary, composition bias did not appear to be a factor in the mt amino acid analysis, as all analyses but one recovered the same Ailuridae-basal topology (see also Arnason et

al., 2007; Yonezawa et al., 2007 Fig. S2A) with generally similar support (Figure 3-4). Although no test for amino acid composition bias is readily available, logdet distance has been shown to be more robust to composition biases than other reconstruction methods and as no difference in result is observed, no compositional bias across taxa is inferred. The one analysis that did not recover the same Ailuridae-basal topology was ME using logdet distance with 4 rate heterogeneity categories, as opposed to 1. In this analysis, Ailuridae and Mephitidae were supported as sister (ME BP 73%), as in the mtDNA analyses.

Rate differences between lineages can also result in long-branch artifacts, as quickly-evolving taxa will accumulate mutations more quickly than other lineages, possibly leading to increased homoplasy. Any rate changes across the tree may have an effect on accurate reconstruction. In this dataset, all but three loci illustrated significant rate variation between some lineages. Of these, all but two loci had significant differences between at least one pair of musteloid lineages. In particular, *ADORA3* and *BRCA1* for Ailuridae, and *APOB* for both Ailuridae and Mephitidae exhibited strong patterns of rate variation compared to other musteloid taxa. However, when these loci were excluded from analysis, a slight decrease in ME and ML BP support was observed, but overall, excluding these loci had little effect. The mtDNA excluding 3$^{rd}$ positions showed the greatest level of rate variation across lineages (Table 3-6). No relationship between the level of rate variation and the recovered topology or its associated support was observed (Table 3-6). Thus, it does not appear that rate variation between lineages is leading to the artifactual recovery of a single topology and yields little insight into which topologies are artifactual and which is "correct".

Thus, it seems that although numerous data biases that can lead to artifactual resolution, including saturation, base composition bias across taxa, and rate differences between lineages, can be identified in the dataset, there is no indication that any one topology (Figure 3-1) is solely the result of artifactual resolution.

*Utility of mtDNA in resolving basal musteloid relationships*

As described above, mtDNA exhibited several biases known to adversely affect phylogenetic reconstruction, including base composition bias and significant rate variation between musteloid lineages. The presence of these biases does not necessarily indicate artifactual resolution, but MP analysis of all mt genes recovers a clearly artifactual grouping of musteloid paraphyly. As this artifactual Ailuridae-Mephitidae+Ursidae grouping is no longer recovered

when mt 3rd codon positions are excluded, either significant base composition bias across taxa or saturation was likely a strong contributer to this problem. However, even mt 1st and 2nd positions combined alone still illustrate saturation (Figure 3-2), and potentially base composition bias across taxa, since logdet analysis of the mt dataset excluding 3rd codon positions results in the recovery of the Mephitidae-basal topology, as opposed to the Ailuridae+Mephitidae sister grouping. More disconcerting is the strongly supported discord between mtDNA and mt amino acid analyses (Table 3-6, Figures 3-3, 3-4). MtDNA analyses generally recover Ailuridae and Mephitidae as sister, although support varies between analyses and with taxon sampling (Tables 3-8, 3-9). Mt amino acid analyses always supported Ailuridae as the basal musteloid lineage, except when logdet analysis with 4 rate heterogeneity categories was applied. One molecule cannot have experienced different evolutionary histories, thus one or both of the recovered topologies must be artifactual. However, both disagree with the topology principally recovered by nuclear analyses. Without knowing the true tree it cannot be determined whether DNA or amino acid sequence is yielding the most appropriate answer for mt evolution within Musteloidea, but it is certain that base composition, saturation, and rate heterogeneity, even at the amino acid level are having an impact on estimating the answer.

When the nuclear dataset is combined with either the mtDNA (excluding 3rd positions) or mt amino acid datasets, the Mephitidae-basal topology recovered by nuclear DNA alone was recovered, but support was drastically reduced from 98% MP BP support for nuclear alone to 62% and 60% respectively for nuclear plus mtDNA or mt amino acid. The nuclear signal appears to "overpower" the mt signal, despite near-equal partition sizes, as nuclear DNA comprises 48% of the nuclear+mtDNA (no 3rd positions) dataset set and 65% of the nuclear+mt amino acid dataset. The strongly conflicting signals between the combined nuclear and the mt datasets are reflected in the poor support when analyzed jointly. Yonezawa et al. (2007) also recovered Mephitidae-basal when using likelihood to analyze nuclear and mt data together, although they recovered almost equal support for an Ailuridae-Mephitidae sister group and neither clade was supported >50% ML BP. They recovered virtually no support for Ailuridae-basal using their TotalML method, although this topology cannot be statistically rejected using the nuclear dataset here (Table 3-7).

It seems reasonable to place more confidence in the combination of multiple, unlinked nuclear loci than in either one of the conflicting mitochondrial answers. As no alternate topologies could be rejected for the mtDNA excluding 3rd positions (Table 3-6), despite strong opposing bootstrap support, mtDNA may not be incongruent with any answer obtained from the

nuclear analyses. What can be determined is that many difficult-to-model processes have occurred during mt evolution within Musteloidea that may have rendered confident resolution of the mt topology impossible at present.

*Individual nuclear genes: support and conflict*

As no single identified data bias is leading consistently to one topology and taxon sampling had little effect, provided that at least two mephitids and one arctoid outgroup are included, the lack of resolution from many genes and the discordance among others may be real. Deep coalescent events may explain the strongly supported discordance between the Mephitidae-basal topology and the Ailuridae-Mephitidae sister grouping (Tables 3-6, 3-7). Most other loci may be lacking sufficient phylogenetic information to resolve this section of the phylogeny. Both ME and MP analyses were generally unable to support resolution for any single gene tree or for the concatenated dataset, indicating a lack of phylogenetic information, as opposed to strong conflict from the separate genes. Because both methods require concatenation of all genes without the ability to model each gene separately, the variability in evolutionary processes across loci may be a significant factor in the inability of these methods to recover and support a topology for the concatenated dataset. This dataset exhibits base composition bias across taxa when concatenated, rate heterogeneity across sites and lineages, high variation in evolutionary rate across loci, and possibly conflicting signal. As such, the concatenated dataset may be too evolutionarily complex for either of these methods or, as with the individual gene trees, there simply may not be enough phylogenetic signal on which to recover relationships.

ML analyses, in contrast, did show conflict between genes, although many genes did not show significant support for any one topology over the other two. The combined dataset (partitioned by gene) strongly supported Ailuridae as the second branch (ML BP=98%), though individual genes were not as clear. From a "democratic vote" standpoint, ML analysis of five loci individually recovered the Mephitidae-basal topology; Ailuridae-basal and Ailuridae-Mephitidae were each recovered by four loci (Table 3-6). However, strong ML BP support was only observed for Mephitidae-basal and Mephitidae-Ailuridae as sister (Table 3-6). For each individual gene, the three proposed topologies (Figure 3-1) were tested to see if any topology could be significantly rejected as an equally good explanation of the data using the AU (Shimodaira, 2002) and SH (Shimodaira and Hasegawa, 1999) tests. More significant results were observed using the AU test than the SH test, as expected, given that AU test is not as

conservative as the SH test (Shimodaira, 2002). The Ailuridae-Mephitidae sister relationship was significantly rejected ($p<0.05$ in the AU test) for the locus *FES*, both Mephitidae-basal and Ailuridae-basal were rejected for *PLCB4*, and Ailuridae-basal was rejected for *RHO* (Table 3-7). When the two most strongly conflicting loci, *FES* and *PLCB4*, were excluded from analysis, little effect was observed in the ML BP result, and the loci appeared to cancel one another out. The Mephitidae-basal topology showed the fewest results that were close to rejection (AU test $p<0.1$), which is reflected in the combined nuclear dataset recovery of this topology. The Ailuridae-Mephitidae sister topology was rejected (AU test $p=0.043$) as an equally probable explanation of the combined nuclear dataset and the Ailuridae-basal topology was close to rejection (AU test $p=0.101$), but the result was not significant, thus, Ailuridae-basal cannot be rejected.

The strong discordance (>90% ML BP support) between *IRBP* and *FES* supporting the Mephitidae-basal topology and *PLCB4* supporting the Ailuridae-Mephitidae sister topology may be the result of deep coalescent events. Approximately half of the genes did not show any strong bootstrap support for a specific placement of Ailuridae within Musteloidea, nor did they statistically reject alternate placements of Ailuridae. It appears that for most nuclear loci, the rapid time in which the lineages radiated did not provide enough time for enough mutations to accumulate and provide phylogenetic signal.

*Mephitidae-basal: Is consistently strong support enough?*

Given the uncertainty across genes and methods, is a consistently recovered and supported topology enough to consider the relationships resolved? As mentioned above, ML analyses of the combined nuclear genes show strong bootstrap support (98%, Figure 3-5) for Ailuridae as the second musteloid branch (the Mephitidae-basal topology). The same ML result was observed whether the loci exhibiting rate variation, significant rejection of alternate topologies, or simply showing the strongest conflict with another locus were excluded. For this particular dataset, the ability to apply different evolutionary models to very differently evolving genes in the ML framework appears to be a strong advantage over ME and MP analyses that inherently do not have this ability, as neither ME or MP could confidently resolve musteloid relationships. ML analysis with the nuclear dataset as a single partition resulted in a decrease in bootstrap support (from 98% to 91%), indicating the importance of separately modeling the differently evolving nuclear loci.

But strong support in phylogenomic studies does not necessarily indicate the true answer, as large gene concatenations can sometimes be positively misleading when short internal branches occur (Carstens and Knowles, 2007; Kubatko and Degnan, 2007). Conversely, concatenation of multiple independent loci may allow emergence of hidden phylogenetic signal (Gatesy and Baker, 2005). If this rapid radiation does show short enough internal branch lengths to place the topology within the 'anomaly zone', anomalous gene trees may be preferentially recovered over those that match the true species tree (Degnan and Rosenberg, 2006). Although it cannot be determined which topology is the true tree, if some loci are recovering anomalous gene trees, there is a statistical preference for balanced topologies, such as the Ailuridae-Mephitidae sister topology over unbalanced topologies, such as Ailuridae- or Mephitidae-basal. Without knowing the effective population size of the lineages at the time of the split, this cannot be definitively illustrated. Recent coalescent-based approaches to reconstructing species trees from gene trees (Carstens and Knowles, 2007; Liu and Pearl, 2007) may be a promising answer to resolving this phylogeny.

Does the evidence provided by nuclear genes, taken in its entirety, provide strong enough evidence for confident resolution of Musteloidea as the Mephitidae-basal topology? Individual genes show conflict, a variety of data biases can be identified, and different methods recover different topologies when the assumptions of their underlying models may be violated. Including more loci that do not show as much bias and/or more mephitid taxa, including the last mephitid genus (*Mydaus*), may yield a more satisfactory answer. Taken together, the strong support from the concatenated dataset, statistical rejection of one alternate topology, and majority 'preference' by individual genes indicate that the Mephitidae-basal topology may represent the best hypothesis of early musteloid evolution (Figure 3-5). This probable rapid radiation appears to have been so difficult to resolve primarily because so little phylogenetic signal exists from the short time between lineage divergences, compounded by possible conflict due to deep coalescent events. Appropriate evolutionary modeling to account for any "non-phylogenetic" signal may allow the true evolutionary signal to be determined. However, until morphological study or fossil discoveries can contribute to the problem and other hypotheses can be definitively excluded or proven to be artifactual, the position of the red panda will remain somewhat enigmatic.

**Table 3-1.** Classification of taxa included in study.

| Classification | Scientific name | Common name |
|---|---|---|
| Carnivora | | |
| Caniformia | | |
| Arctoidea | | |
| Musteloidea | | |
| Ailuridae | | |
| | *Ailurus fulgens* | Red panda |
| Mephitidae | | |
| | *Mephitis mephitis* | Striped skunk |
| | *Spilogale gracilis* | Western spotted skunk |
| | *Conepatus mesoleucus* | Hog-nosed skunk |
| Procyonidae | | |
| | *Procyon lotor* | North American raccoon |
| | *Nasua narica* | White-nosed coati |
| | *Bassariscus astutus* | Ringtail |
| | *Potos flavus* | Kinkajou |
| Mustelidae | | |
| | *Taxidea taxus* | American badger |
| | *Meles meles* | European badger |
| | *Gulo gulo* | Wolverine |
| | *Martes americana* | American marten |
| | *Mustela vison* | American mink |
| | *Mustela putorius furo* | Domestic ferret |
| | *Mustela ermina* | Ermine |
| | *Mustela nivalis* | Least weasel |
| | *Enhydra lutris* | Sea otter |
| | *Lontra canadensis* | North American river otter |
| Pinnipedia | | |
| Phocidae | | |
| | *Erignathus barbatus* | Bearded seal |
| | *Phoca vitulina* | Harbour seal |
| | *Mirounga angustirostris* | Northern elephant seal |
| | *Hydrurga leptonyx* | Leopard seal |
| Odobenidae | | |
| | *Odobenus rosmarus* | Walrus |
| Otariidae | | |
| | *Eumetopias jubatus* | Steller's sea lion |
| Ursidae | *Ailuropoda melanoleuca* | Giant panda |
| | *Melursus ursinus* | Sloth bear |
| | *Helarctos malayanus* | Sun bear |
| | *Ursus americanus* | American black bear |
| | *Ursus maritimus* | Polar bear |
| Canidae | | |
| | *Canis lupus* | Wolf |
| | *Alopex lagopus* | Arctic fox |
| Feliformia | | |
| Felidae | | |
| | *Felis catus* | Domestic cat |
| | *Lynx canadensis* | Canadian lynx |

**Table 3-2.** Loci included in this study.

| Gene | Acronym | Fragment | Primer Reference | Forward Primer / Reverse Primer |
|---|---|---|---|---|
| Adenosine A3 receptor | *ADORA3* | exon 2 | Murphy et al., 2001 | ACCCCCATGTTTGGCTGGAA |
| | | | | GATAGGGTTCATCATGGAGTT |
| Apolipoprotein B | *APOB* | exon 26 | Jiang et al., 1998 (J1R) | GTGCCAGGTTCAATCAGTATAAGT |
| | | | Amrine-Madsen et al., 2003 (187F) | CCAGCAAAATTTTCTTTTACTTCAA |
| Brain-derived neurotrophic factor | *BDNF* | exon 1 | Murphy et al., 2001 | CATCCTTTTCCTTACTATGGTT |
| | | | | TTCCAGTGCCTTTTGTCTATG |
| Breast cancer 1, early onset | *BRCA1* | exon 9 (fragment 1) | Lindblad-Toh et al., 2005 | GCC ATG TGG CAC ARA TRC TC |
| | | | | CTC TRC TTT CTT GAT AAA RTC CTC AG |
| Interphotoreceptor binding protein | *IRBP* | exon 1 | Stanhope et al., 1992 | ATGGCCAAGGTCCTCTTGGATAACTACTGCTT |
| | | | (+217,-1531) | CGCAGGTCCATGATGAGGTGCTCCGTGTCCTG |
| | | | Fulton & Strobeck 2006 | GGTACAGTGCCGACAAAGATG |
| | | | (internal +785,-913) | GCTTCTGGAGGTCCAGGGC |
| Prepronociceptin | *PNOC* | exon 2 | Murphy et al., 2001 | GCATCCTTGAGTGTGAAGAGAA |
| | | | | TGCCTCATAAACTCACTGAACC |
| Recombination activating gene 1 | *RAG1* | exon 1 | Teeling et al., 2000 | GCTTTGATGGACATGGAAGAAGACAT |
| | | | (F1705,R2864) | GAGCCATCCCTATCAATAATTTCAGG |
| | | | Sato et al. 2004 | AGCCTCCCAAAATCTTGTCTTCCACTCCA |
| | | | (internal F2357,R2486) | AATGTCACAGTGAAGGGCATCTATGGAAGG |
| Recombination activating gene 2 | *RAG2* | exon 1 | Murphy et al., 2001 | TCATGGAGGGAAAACACCAAA |
| | | | | TGCACTGGAGACAGAGATTC |
| Cholinergic receptor, nicotinic, alpha 1 | *CHRNA1* | intron 8 | Lyons et al., 1997 | GACCATGAAGTCAGACCAGGAG |
| | | | | GGAGTATGTGGTCCATCACCAT |

**Table 3-2.** *Continued.*

| Gene | Acronym | Fragment | Primer Reference | Forward Primer / Reverse Primer |
|---|---|---|---|---|
| Feline sarcoma oncogene | *FES* | intron 14 | Venta et al., 1996 | GGGGAACTTTGGCGAAGTGTT |
| | | | | TCCATGACGATGTAGATGGG |
| Feline leukemia virus subgroup C cellular receptor 1 | *FLVCR1* | intron 7 | Housley et al., 2006 | TTGAAATCACTTACCCTGAATCTGA |
| | | | | TCCTTGAGCCAATGTGAACA |
| Growth hormone receptor | *GHR* | intron 9 | Venta et al., 1996 | CCAGTTCCAGTTCCAAAGAT |
| | | | | TGATTCTTCTGGTCAAGGCA |
| Rhodopsin | *RHO* | intron 3 | Venta et al., 1996 | TACATGTTCGTGGTCCACTT |
| | | | | TGGTGGGTGAAGATGTAGAA |
| Amyloid beta (A4) precursor protein | *APP* | 3' UTR | Murphy et al., 2001 | TCCAAGATGCAGCAGAACG |
| | | | | CTAATGTGTGCACATAAAACAGG |
| 12 mitochondrial protein-coding genes | mtDNA | mitochondrion | Delisle & Strobeck 2002 | |

**Table 3-3.** Nuclear loci accession numbers and references. Information for the five genes used in Chapter 2 is listed in Table 2-1. **This study, available upon request [a]Koepfli et al. (2007) [b]Sato et al. (2006) [c]Sato et al. (2004) [d]Yonezawa et al. (2007) [e]Johnson et al. (2006) [f]Lindblah-Toh et al. (2005) [g]Lin (unpublished) [h]Feng (unpublished) [i]Madsen et al. (2001)

| Species | *ADORA3* | *APOB* | *RAG1* | *RAG2* | *BDNF* | *PNOC* | *BRCA1* | *FLVCR1* | *PLCB4* |
|---|---|---|---|---|---|---|---|---|---|
| *U. maritimus* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *U. americanus* | ** | ** | ** | ** | DQ240386[f] | ** | DQ240420[f] | ** | ** |
| *H. malayanus* | ** | ** | | ** | AF002240[g] | ** | ** | ** | ** |
| *M. ursinus* | ** | AB193428[b] | AB109362[c] | ** | ** | ** | ** | ** | ** |
| *A. melanoleuca* | - | - | AB302262[d] | - | U56638[b] | - | DQ240421[f] | - | - |
| *P. lotor* | DQ660175[a] | AB193427[b] | AB109359[c] | DQ660279[a] | AF00318[g] | DQ660252[a] | ** | ** | ** |
| *N. narica* | DQ660171[a] | ** | DQ660262[a] | DQ660275[a] | ** | DQ660248[a] | ** | ** | ** |
| *B. astutus* | DQ660169[a] | ** | DQ660260[a] | DQ660274[a] | ** | DQ660246[a] | ** | ** | ** |
| *P. flavus* | DQ660173[a] | ** | ** | DQ660277[a] | ** | DQ660250[a] | - | ** | ** |
| *A. fulgens* | ** | AB193430[b] | AB188525[b] | ** | U56639[b] | ** | ** | ** | ** |
| *M. mephitis* | ** | AB193406[b] | AB109358[c] | ** | ** | ** | ** | ** | ** |
| *S. gracilis* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *C. mesoleucus* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *E. lutris* | ** | AB193403[b] | ** | DQ660280[a] | ** | DQ660253[a] | ** | ** | ** |
| *L. canadensis* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *M. americana* | DQ660179[a] | AB193408[b] | AB109341[c] | DQ660283[a] | ** | DQ660256[a] | ** | ** | ** |
| *T. taxus* | DQ660180[a] | DQ660194[a] | ** | DQ660284[a] | ** | DQ660257[a] | ** | ** | ** |
| *M. meles* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *G. gulo* | ** | AB193407[b] | AB109340[c] | ** | ** | ** | ** | ** | ** |
| *M. vison* | DQ660177[a] | AB193421[b] | AB109354[c] | DQ660281[a] | ** | DQ660254[a] | ** | ** | ** |
| *M. putorius* | ** | AB193418[b] | AB109351[c] | ** | ** | ** | ** | ** | ** |
| *M. ermina* | ** | AB193414[b] | AB109347[c] | ** | ** | ** | ** | ** | ** |
| *M. nivalis* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *E. jubatus* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *O. rosmarus* | ** | ** | ** | ** | ** | ** | DQ240423[f] | ** | ** |
| *E. barbatus* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *P. vitulina* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *H. leptonyx* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *M. angustirostris* | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| *C. lupus* | ** | ** | ** | - | ** | ** | DQ240395[f] | ** | ** |
| *A. lagopus* | ** | ** | ** | - | ** | ** | DQ240390[f] | ** | ** |
| *F. catus* | ** | ** | ** | DQ082330[e] | ** | DQ082285[e] | AF284018[i] | ** | ** |
| *L. canadensis* | ** | ** | ** | DQ082347[e] | ** | DQ082302[e] | ** | ** | ** |

**Table 3-4.** Mitochondrial gene accession numbers and references. In some cases, sequence for *cytochrome b* was obtained from a different source. *Individual gene accessions in Table 5-5.

| Species | GenBank # | Reference | *Cytochrome b* | |
|---|---|---|---|---|
| *U. maritimus* | NC_003428 | Delisle & Strobeck (2002) | | |
| *U. americanus* | NC_003426 | Delisle & Strobeck (2002) | | |
| *H. malayanus* | AY598584-94 | Delisle & Strobeck (2005) | U18899 | Talbot & Shields (1996) |
| *M. ursinus* | ** | This study | U23560 | Talbot & Shields (1996) |
| *A. melanoleuca* | AM711896 | Arnason et al. (2007) | | |
| *P. lotor* | AY598573-83 | Delisle & Strobeck (2005) | X94930 | Ledje & Arnason (1996) |
| *N. narica* | ** | This study | DQ533940 | Fulton & Strobeck (2007) |
| | DQ533934, 944 | Fulton & Strobeck (2007) | | |
| *B. astutus* | ** | This study | AF498159 | Koepfli & Wayne (2003) |
| | DQ533935, 945 | Fulton & Strobeck (2007) | | |
| *P. flavus* | ** | This study | DQ660304 | Koepfli et al. (2007) |
| | DQ533936, 946 | Fulton & Strobeck (2007) | | |
| *A. fulgens* | AY59818-28 | Delisle & Strobeck (2005) | X94919 | Ledje & Arnason (1996) |
| *M. mephitis* | AY598529-39 | Delisle & Strobeck (2005) | X94927 | Ledje & Arnason (1996) |
| *S. gracilis* | ** | This study | | |
| *C. mesoleucus* | ** | This study | | |
| *E. lutris* | ** | This study | DQ533942 | Fulton & Strobeck (2007) |
| | DQ533938, 948 | Fulton & Strobeck (2007) | | |
| *L. canadensis* | AY59851-61 | Delisle & Strobeck (2005) | AF057121 | Koepfli & Wayne (1998) |
| *M. americana* | AY598540-50 | Delisle & Strobeck (2005) | AY121352 | Stone et al. (2002) |
| *T. taxus* | AY59862-72 | Delisle & Strobeck (2005) | AF057132 | Koepfli & Wayne (1998) |
| *M. meles* | ** | This study | X94922 | Ledje & Arnason (1996a) |
| *G. gulo* | AY377143-384 | Davis et al. (2004) | AB051245 | Hosoda et al. (2000) |
| *M. vison* | AY377143-384 | Davis et al. (2004) | AB026109 | Kurose et al. (2000) |
| *M. putorius* | ** | This study | | |
| *M. ermina* | ** | This study | | |
| *M. nivalis* | ** | This study | DQ533943 | Fulton & Strobeck (2007) |
| | DQ533939, 949 | Fulton & Strobeck (2007) | | |
| *E. jubatus* | NC_004030 | Arnason et al. (2006) | | |
| *O. rosmarus* | AY377143-384* | Davis et al. (2004) | ** | This study |
| *E. barbatus* | AY377143-384* | Davis et al. (2004) | AY140982 | Palo & Vainola (2006) |
| *P. vitulina* | AM181032 | Arnason et al. (2006) | | |
| *H. leptonyx* | AY377143-384* | Davis et al. (2004) | | |
| *M. angustirostris* | AY377143-384* | Davis et al. (2004) | | |
| *C. lupus* | AY598494-505 | Delisle & Strobeck (2005) | | |
| *A. lagopus* | AY598506-517 | Delisle & Strobeck (2005) | | |
| *F. catus* | NC_001700 | Lopez et al. (1996) | | |
| *L. canadensis* | AY598470-481 | Delisle & Strobeck (2005) | | |

**Table 3-5.** Significant relative rate comparisons between lineages (GRate). Each nuclear gene was examined separately; mtDNA was examined as a single partition, excluding 3rd codon position bases. Members of Felidae were used as the outgroup. Loci discussed in text are highlighted in bold. Significance levels: * = <0.05, ** = <0.01, *** = <0.0001, **** = <0.0001.

| | Mephitidae | Procyonidae | Mustelidae | Pinnipedia | Ursidae | Canidae |
|---|---|---|---|---|---|---|
| Ailuridae | **ADORA3**** | **ADORA3**** | **ADORA3*** | **ADORA3*** | **ADORA3**** | FES**** |
| | **BRCA1**** | **BRCA1*** | **BRCA1***** | CHRNA1*** | GHR** | IRBP* |
| | PLCB4** | **APOB**** | FLVCR1** | **APOB*** | **APOB*** | |
| | FLVCR* | | FES* | FES* | FES* | |
| | RHO* | | | | | |
| | **mtDNA**** | **mtDNA*** | **mtDNA*** | | | **mtDNA**** |
| Mephitidae | - | **APOB**** | PLCB4* | **APOB*** | **APOB**** | FES**** |
| | | | RHO** | RHO*** | GHR*** | RHO** |
| | | | | IRBP** | IRBP** | FLVCR1* |
| | | | | CHRNA1** | | |
| | | | | **BRCA1**** | | |
| | | | | **ADORA3*** | | |
| | | | | GHR* | | |
| | | | | FES* | | |
| | | **mtDNA*** | **mtDNA*** | **mtDNA***** | **mtDNA**** | |
| Procyonidae | | - | FLVCR1* | CHRNA1*** | GHR** | FES**** |
| | | | **APOB*** | **BRCA1**** | FES** | |
| | | | | FES** | CHRNA1* | |
| | | | | **ADORA3*** | IRBP* | |
| | | | | RHO* | | |
| | | | | IRBP* | | |
| Mustelidae | | | - | FES**** | FES*** | FES**** |
| | | | | **BRCA1****** | GHR** | |
| | | | | CHRNA1*** | CHRNA1* | |
| | | | | FLVCR1* | | FLVCR1***** |
| Pinnipedia | | | | - | BRCA1* | FES**** |
| | | | | | **ADORA3*** | IRBP*** |
| Ursidae | | | | | - | FES**** |
| | | | | | | IRBP***** |
| | | | | | | **mtDNA*** |

**Table 3-6.** Locus information, ML results, and relative rate comparisons. Maximum likelihood (ML) topology and bootstrap support (BP) determined using RAxML. Strong (>75%) BP values are in bold. The number of significant relative rate differences between musteloid lineages was calculated in GRate. bp = base pairs. PI = Parsimony Informative. **FLVCR1* supports Ailuridae-basal, but also Mephitidae+Procyonidae (ML BP=90).

| | Model | Length (bp) | % PI sites | ML BP Fig. 3-1a | ML BP Fig. 3-1b | ML BP Fig. 3-1c | # signif. rate differences in Musteloidea |
|---|---|---|---|---|---|---|---|
| *FES* | GTR+G | 445 | 36.0 | **99** | | | 1 |
| *IRBP* | HKY+I+G | 1177 | 18.7 | **96** | | | 1 |
| *PLCB4* | GTR+G | 327 | 29.7 | | | **96** | 2 |
| *RHO* | HKY+G | 277 | 27.4 | | | **83** | 2 |
| *BRCA1* | GTR+G | 689 | 26.1 | **78** | | | 3 |
| *FLVCR1* | GTR+G | 282 | 39.4 | | 74* | | 2 |
| *ADORA3* | GTR+G | 354 | 23.4 | 69 | | | 3 |
| *CHRNA1* | SYM+G | 371 | 38.8 | | 65 | | 0 |
| *GHR* | GTR+G | 636 | 29.2 | <50 | | | 0 |
| *BDNF* | GTR+I+G | 566 | 11.1 | | | <50 | 0 |
| *PNOC* | HKY+G | 268 | 20.9 | | | 50 | 0 |
| *RAG2* | GTR+G | 444 | 12.6 | | <50 | | 0 |
| *APOB* | GTR+G | 940 | 19.3 | | <50 | | 3 |
| All nuclear | GTR+I+G | 6776 | 23.8 | **98** | | | - |
| mtDNA | GTR+I+G | 10842 | 46.7 | | | **83** | - |
| mtDNA no 3rd pos | GTR+I+G | | | | | 53 | 5 |
| mt protein | MTMAM | 3619aa | 25.9 | | **91** | | - |

**Table 3-7.** Hypothesis tests. The favoured tree (the ML tree, except for *GHR*) is highlighted in bold. Probability values less than 0.1 are underlined and those less than 0.05 are starred.

| | AU test: Probability (Standard Error) | | | SH test: Probability (Standard Error) | | |
|---|---|---|---|---|---|---|
| **Locus** | **Meph-basal** | **Ailur-basal** | **Sister** | **Meph-basal** | **Ailur-basal** | **Sister** |
| *IRBP* | **0.926 (0.004)** | 0.128 (0.007) | 0.067 (0.006) | **0.901 (0.003)** | 0.116 (0.003) | 0.101 (0.003) |
| *FES* | **0.958 (0.003)** | 0.069 (0.005) | 0.021* (0.003) | **0.933 (0.003)** | 0.08 (0.003) | 0.063 (0.002) |
| *GHR* | **0.595 (0.01)** | 0.579 (0.01) | 0.268 (0.009) | 0.632 (0.005) | **0.652 (0.005)** | 0.623 (0.005) |
| *CHRNA1* | 0.116 (0.006) | **0.792 (0.007)** | 0.369 (0.01) | 0.273 (0.004) | **0.765 (0.004)** | 0.309 (0.005) |
| *BRCA1* | **0.903 (0.005)** | 0.057 (0.004) | 0.151 (0.006) | **0.853 (0.004)** | 0.149 (0.004) | 0.159 (0.004) |
| *BDNF* | 0.21 (0.007) | **0.906 (0.004)** | 0.094 (0.004) | 0.216 (0.004) | **0.791 (0.004)** | 0.206 (0.004) |
| *PNOC* | 0.23 (0.008) | 0.23 (0.008) | **0.77 (0.008)** | 0.272 (0.004) | 0.272 (0.004) | **0.728 (0.004)** |
| *RHO* | 0.36 (0.01) | 0.044* (0.005) | **0.735 (0.008)** | 0.342 (0.005) | 0.188 (0.004) | **0.808 (0.004)** |
| *RAG2* | 0.273 (0.008) | **0.877 (0.005)** | 0.094 (0.004) | 0.224 (0.004) | **0.79 (0.004)** | 0.215 (0.004) |
| *ADORA3* | **0.854 (0.006)** | 0.146 (0.006) | 0.146 (0.006) | **0.754 (0.004)** | 0.246 (0.004) | 0.246 (0.004) |
| *APOB* | 0.281 (0.008) | 0.282 (0.008) | **0.721 (0.008)** | 0.299 (0.005) | 0.299 (0.005) | **0.701 (0.005)** |
| *PLCB4* | 0.034* (0.003) | 0.034* (0.003) | **0.966 (0.003)** | 0.083 (0.003) | 0.083 (0.003) | **0.917 (0.003)** |
| *FLVCR1* | 0.098 (0.005) | **0.902 (0.005)** | 0.099 (0.005) | 0.175 (0.004) | **0.825 (0.004)** | 0.175 (0.004) |
| All nuclear | **0.940 (0.004)** | 0.101 (0.006) | 0.043* (0.005) | **0.944 (0.002)** | 0.091 (0.003) | 0.054 (0.002) |
| mtnopos3 | 0.175 (0.008) | 0.465 (0.010) | **0.671 (0.009)** | 0.320 (0.005) | 0.541 (0.005) | **0.755 (0.004)** |

**Table 3-8.** Effects of Mephitidae taxon sampling on the complete mtDNA result. Sister = Ailuridae + Mephitidae. * = Ailuridae + Mephitidae recovered as sister to Ursidae, not Musteloidea. ME = minimum evolution (logdet), MP = maximum parsimony, ML = maximum likelihood, BP = bootstrap support.

| Included Mephitidae genera | ME BP | MP BP | ML BP |
|---|---|---|---|
| *Mephitis* | 35% sister | 96% sister* | 93% sister |
| *Conepatus* | 77% sister | 92% sister* | 95% sister |
| *Spilogale* | 50% Ailuridae-basal | 89% sister | 85% sister |
| *Mephitis, Conepatus* | 65% sister | 74% sister* | 76% sister |
| *Mephitis, Spilogale* | 44% Ailuridae-basal | 79% sister* | 57% sister |
| *Conepatus, Spilogale* | 55% sister | 68% sister* | 78% sister |
| *Mephitis, Conepatus, Spilogale* | 55% sister | 59% sister* | 78% sister |

**Table 3-9.** Effects of outgroup selection on nuclear results.

| Non-musteloids included: | ME BP (logdet) | MP BP | ML BP |
|---|---|---|---|
| bears | 83% Mephitidae-basal | 93% Mephitidae-basal | 98% Mephitidae-basal |
| pinnipeds | 57% sister | 54% Mephitidae-basal | 95% Mephitidae-basal |
| bears, pinnipeds | 65% Mephitidae-basal | 48% Mephitidae-basal | 100% Mephitidae-basal |
| bears, pinnipeds, canids | 68% Mephitidae-basal | 55% sister | 93% Mephitidae-basal |
| felids | 68% sister | 72% Mephitidae-basal | 41% Mephitidae-basal |
| canids | 63% Mephitidae-basal | 64% Mephitidae-basal | 59% Mephitidae-basal |
| All | 56% Mephitidae-basal | 50% sister | 98% Mephitidae-basal |

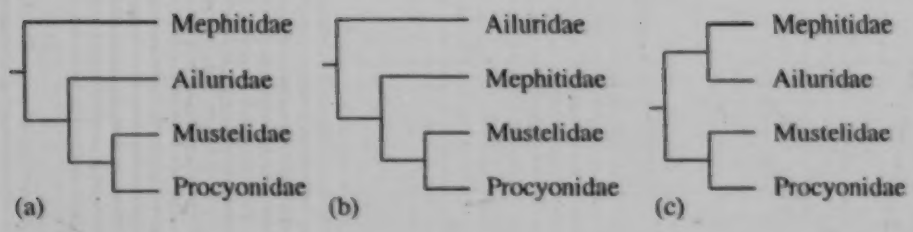


**Figure 3-1.** Hypothesized relationships between musteloid families. (a) Mephitidae-basal (b) Ailuridae-basal (c) Ailuridae-Mephitidae sister.

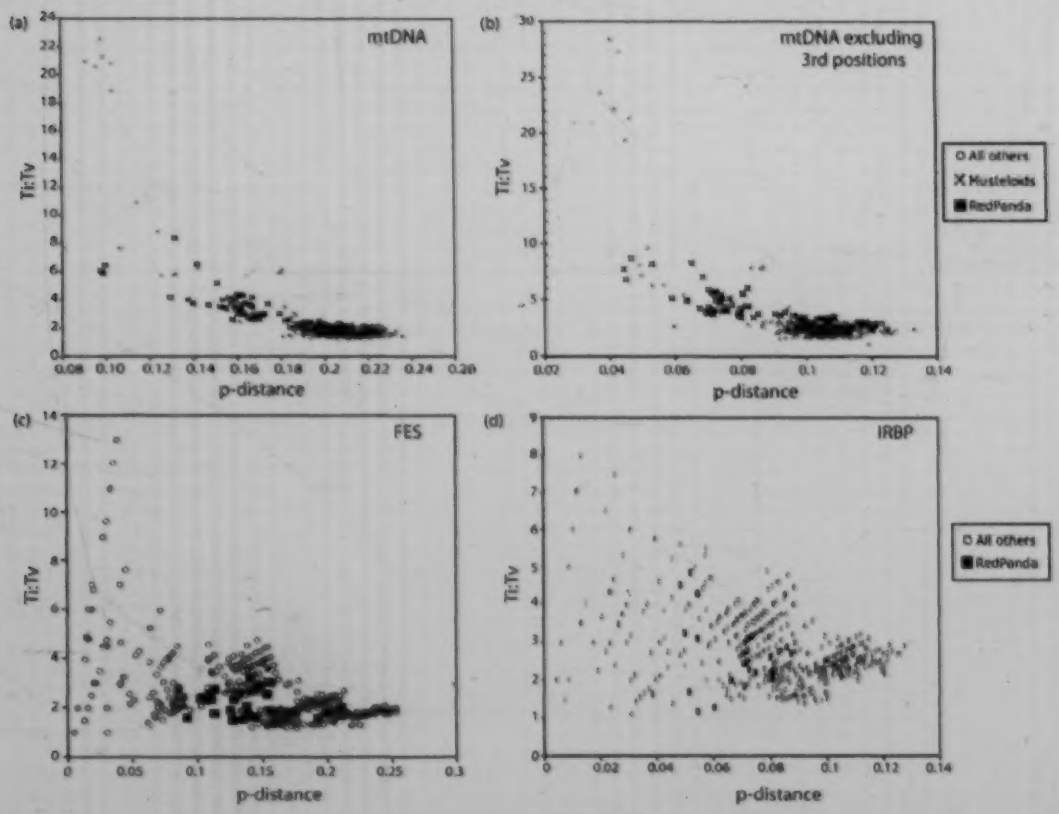


**Figure 3-2.** Saturation plots. The transition-transversion (Ti:Tv) ratio is plotted against uncorrected p-distance for each pair-wise taxa comparison, as calculated in PAUP*. (a) mtDNA (b) mtDNA with 3rd codon positions excluded (c) *FES* (d) *IRBP*. Grey circles represent outgroup pair-wise comparisons, x's represent comparisons within Musteloidea not including the red panda (these are included in the circle group for the nuclear loci), black squares represent any pair-wise comparison including the red panda.

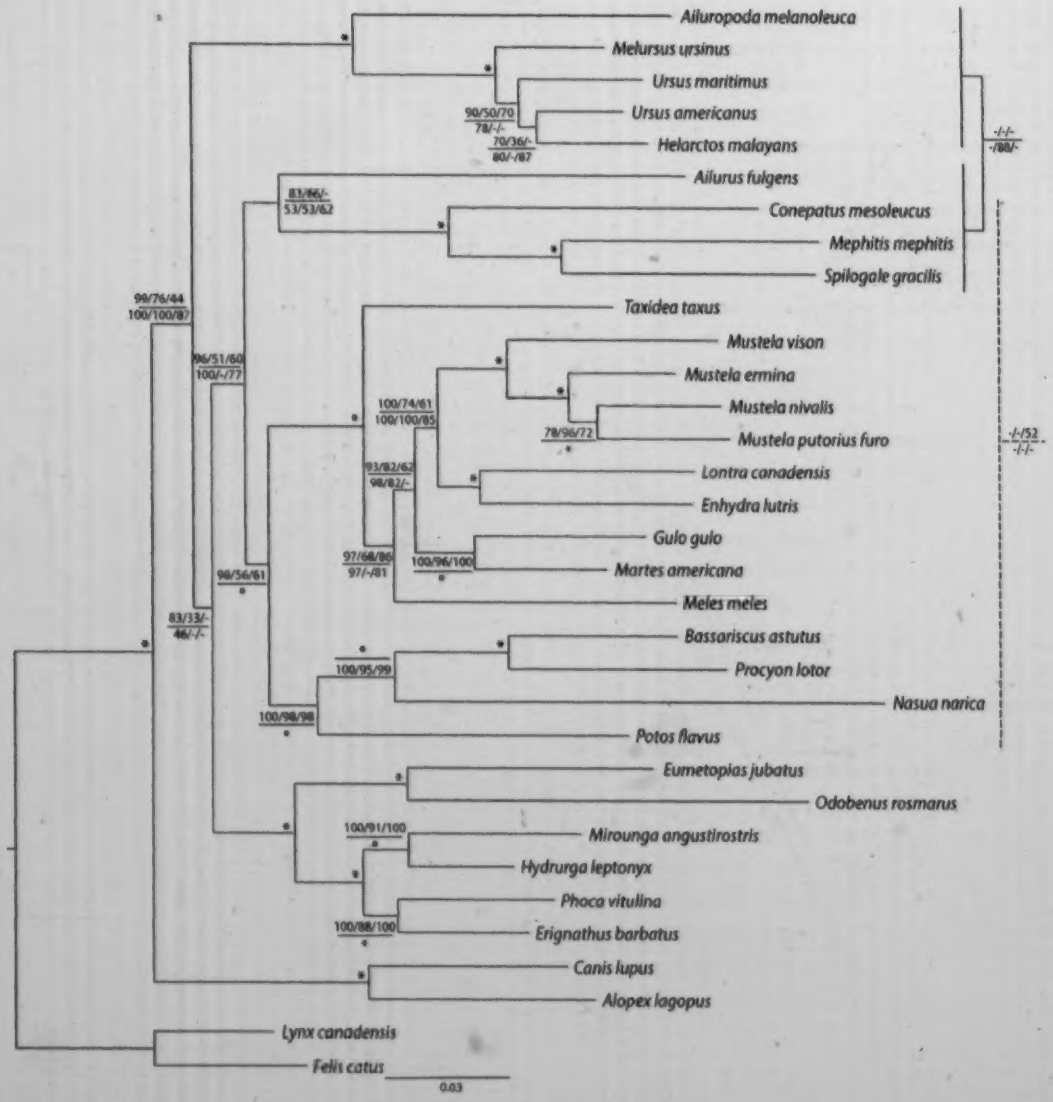


**Figure 3-3.** Maximum likelihood topology for MtDNA excluding 3rd codon positions. Topology and branch lengths were estimated using RAxML. Support values represent MLBP/MPBP/MEBP for mtDNA excluding 3rd positions above the same values for the complete mtDNA, partitioned by gene and codon position for the ML analysis. * = ≥99% BP support from all methods. - = node not recovered by that analysis. Alternate topologies are indicated on the right.

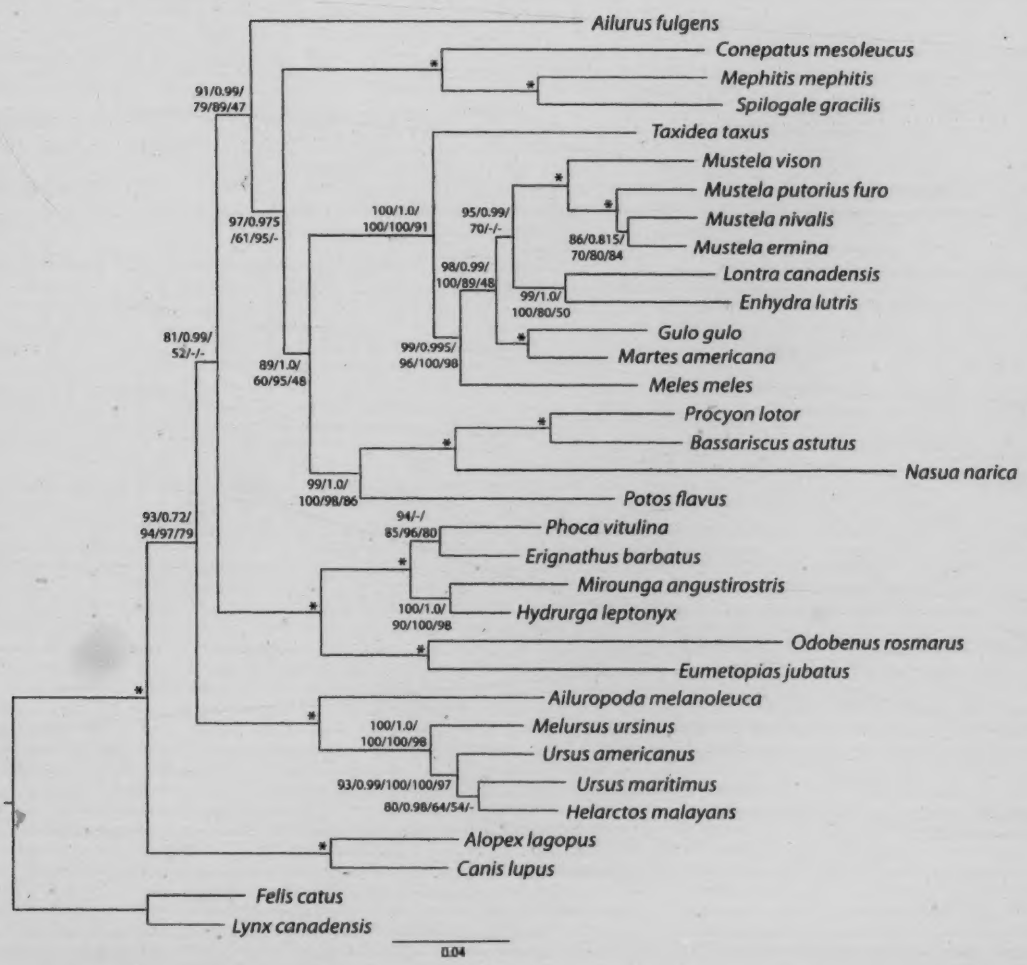


**Figure 3-4.** Mt amino acid maximum likelihood topology. Topology and branch lengths were estimated in RAxML using the MTMAM-F model. The same topology was recovered using Bayesian MCMC estimation in Phylobayes, maximum parsimony (MP) analysis, and minimum evolution (ME) using logdet distance with 1 rate heterogeneity category. ME logdet with 4 rate categories recovered Ailuridae and Mephitidae as sister (ME BP =73). Support values correspond to MLBP/BPP/MPBP/MEBP1cat/MEBP4cat. * = nodes supported by ≥99% BP and BPP=1.0 by all methods. - = node not recovered by that analysis.

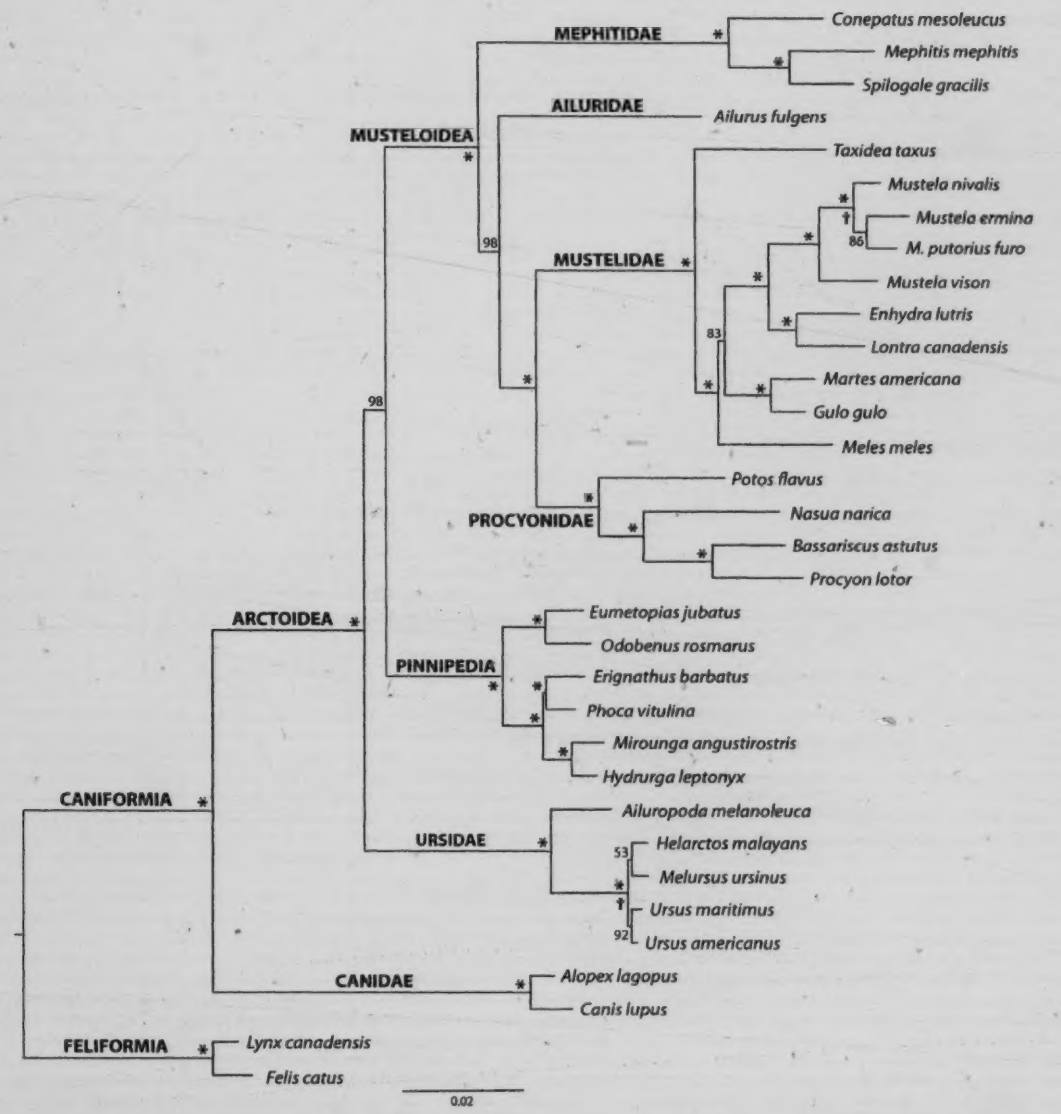


**Figure 3-5.** Nuclear maximum likelihood topology. Topology, branch lengths, and bootstrap support values were determined using RAxML. The same Musteloidea family relationships (Mephitidae-basal) were recovered using minimum evolution with logdet distance (ME), but not with maximum parsimony (MP). Mephitidae-basal was also recovered by ML analysis of both nuclear and mt combined datasets, and ME, but not MP analysis of the nuclear + mt (excluding 3rd codon positions) dataset. † = clades not stably resolved between different analyses of different datasets. *=clades with 100% ML bootstrap support.

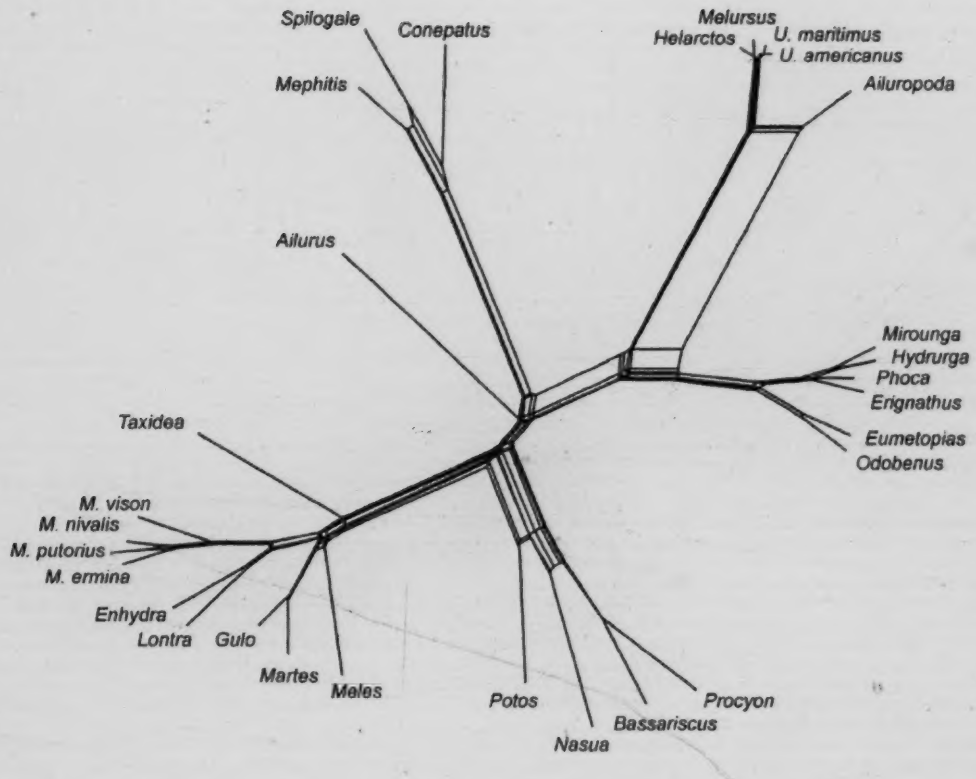


**Figure 3-6.** NeighbourNet analysis of concatenated nuclear loci using LogDet distance.

**References**


Amrine-Madsen, H., Koepfli, K.-P., Wayne, R.K., Springer, M.S., 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. Mol.Phylogenet.Evol. 28, 225-240.

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., 2007. Mitogenomic analyses of caniform relationships. Mol.Phylogenet.Evol. 45, 863-874.

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., Lehman, N., Petrov, E.A., Vainola, R., 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol.Phylogenet.Evol. 41, 345-354.

Barker, F.K., Lutzoni, F.M., 2002. The utility of the incongruence length difference test. Syst.Biol. 51, 625-637.

Baskin, J.A., 1998. Procyonidae. In: Janis, C.M., Scott, K.M., Jacobs, L.J. (Eds.), Evolution of Tertiary mammals of North America. Volume 1: Terrestrial carnivores, ungulates, and ungulatelike mammals. Cambridge University Press, New York, NY, pp. 144-151.

Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst.Biol. 56, 400-411.

Darlu, P., Lecointre, G., 2002. When does the incongruence length difference test fail? Mol.Biol.Evol. 19, 432-437.

Davis, C.S., Delisle, I., Stirling, I., Siniff, D.B., Strobeck, C., 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. Mol.Phylogenet.Evol. 33, 363-377.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics 2, e68.

Delisle, I., Strobeck, C., 2002. Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. Mol.Biol.Evol. 119, 357-361.

Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Mol.Phylogenet.Evol. 37, 192-201.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol.Phylogenet.Evol. 17, 401-406.

Domingo-Roura, X., Lopez-Giraldez, F., Saeki, M., Marmi, J., 2005. Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores. Genet.Res.Cambridge 85, 223-233.

Dowton, M., Austin, A.D., 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy - the behavior of the incongruence length difference test in mixed-model analyses. Syst.Biol. 51, 19-31.

Dragoo, J.W., Honeycutt, R.L., 1997. Systematics of Mustelid-like carnivores. J.Mammal 78, 426-443.

Edwards, S.V., Jennings, W.B., Shedlock, A.M., 2005. Phylogenomics of modern birds in the era of genomics. Proc.Roy.Soc.B 272, 979-992.

Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1995. Constructing a significance test for incongruence. Syst.Biol. 44, 570-572.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst.Zool. 27, 401-410.

Finarelli, J.A., 2008. A total evidence phylogeny of the Arctoidea (Carnivora: Mammalia): relationships among basal taxa. J.Mamm.Evol. in press.

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M.A., 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst.Biol. 54, 317-337.

Flynn, J.J., Nedbal, M.A., Dragoo, J.W., Honeycutt, R.L., 2000. Whence the red panda? Mol.Phylogenet.Evol. 17, 190-199.

Foster, P.G., 2004. Modeling compositional heterogeneity. Syst.Biol. 53, 485-495.

Fulton, T.L., Strobeck, C., 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. Mol.Phylogenet.Evol. 41, 165-181.

Fulton, T.L., Strobeck, C., 2007. Novel phylogeny of the raccoon family (Procyonidae: Carnivora) based on nuclear and mitochondrial DNA evidence. Mol.Phylogenet.Evol. 43, 1171-1177.

Galtier, N., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc.Natl.Acad.Sci. 92, 11317-11321.

Gatesy, J., Baker, R.H., 2005. Hidden likelihood support in genomic data: can forty-five wrongs make a right? Syst.Biol. 54, 483-492.

Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst.Biol. 47, 9-17.

Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J.Syst.Evol. 46, 239-257.

Hosoda, T., Suzuki, H., Harada, M., Tsuchiya, K., Han, S.-H., Zhang, Y.-p., Kryukov, A.P., Lin, L.-K., 2000. Evolutionary trends of the mitochondrial lineage differentiation in species of genera Martes and Mustela. Genes Genet.Syst. 75, 259-267.

Housley, D.J.E., Zalewski, Z.A., Beckett, S.E., Venta, P.J., 2006. Design factors that influence PCR amplification success of cross-species primers among 1147 mammalian primer pairs. BMC Genomics 7, 253.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol.Biol.Evol. 23, 254-267.

Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, A.W.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimates. Syst.Biol. 53, 638-643.

Jiang, Z., Priat, C., Galibert, F., 1998. Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. Mamm.Genome 9, 577-587.

Johnson, W.E., Eizirik, E., Pecon-Slattery, J., Murphy, W.J., Antunes, A., Teeling, E., O'Brien, S.J., 2006. The late Miocene radiation of modern Felidae: a genetic assessment. Science 311, 73-77.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066.

Koepfli, K.-P., Gompper, M.E., Eizirik, E., Ho, C.-C., Linden, L., Maldonado, J.E., Wayne, R.K., 2007. Phylogeny of the Procyonidae (Mammalia: Carnivora): molecules, morphology and the Great American Interchange. Mol.Phylogenet.Evol. 43, 1076-1095.

Koepfli, K.-P., Wayne, R.K., 1998. Phylogenetic relationships of otters (Carnivora: Mustelidae) based on mitochondrial cytochrome b sequences. J.Zool.Lond. 246, 401-416.

Koepfli, K.-P., Wayne, R.K., 2003. Type I STS markers are more informative than cytochrome b in phylogenetic reconstruction of the Mustelidae (Mammalia: Carnivora). Syst.Biol. 52, 571-593.

Kolaczkowski, B., Thornton, J.W., 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogenous. Nature 431, 980-984.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concantenated data under coalescence. Syst.Biol. 56, 17-24.

Kuhner, M.K., Felsenstein, J., 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol.Biol.Evol. 11, 459-468.

Kurose, N., Abramov, A.V., Masuda, R., 2000. Intrageneric diversity of the cytochrome b gene and phylogeny of eurasian species of the genus Mustela (Mustelidae, Carnivora). Zool.Sci. 17, 673-679.

Lartillot, N., Blanquart, S., Lepage, T., 2007. Phylobayes 2.3. http://www.atgc-montpellier.fr/phylobayes/

Lartillot, N., Phillipe, H., 2004. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol.Biol.Evol. 21, 1095-1109.

Ledje, C., Arnason, U., 1996. Phylogenetic analyses of complete cytochrome b genes of the order Carnivora with particular emphasis on the Caniformia. J.Mol.Evol. 42, 135-144.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas III, E.J., Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.A., deJong, P.J., Kirkness, E.F., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.-W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.-P., Parker, H.G., Pollinger, J.P., Searle, S.M.J., Sutter, N., Thomas, R., Webber, C., Platform, B.I.G.S., Lander, E.S., 2005. Genome sequence, comparative analysis and haplotypes structure of the domestic dog. Nature 438, 803-819.

Liu, L., Pearl, D.K., 2007. Species trees from gene trees: Reconstructing posterior distributions of a species phylogeny using estimated gene tree distributions. Syst.Biol. 56, 504-514.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol.Biol.Evol. 11, 605-612.

Lopez, J.V., Cevario, S., O'Brien, S.J., 1996. Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics 33, 229-246.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E., O'Brien, S.J., 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat.Genet. 15, 47-56.

Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary analysis. Version 1.12. http://mesquiteproject.org

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409, 610-614.

McKenna, M.C., Bell, S.K., 1997. Classification of mammals above the species level. Columbia University Press, New York, USA.

Muller, K., 2002. GRate - relative rate tests for taxon groups using general DNA substitution models in PAUP*. v.04. http://www.nees.uni-bonn.de/downloads/GRate/GRate.htm

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614-618.

Nylander, J.A.A., 2004. MrModeltest. Version 2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Appl.Biosc. 12, 357-358.

Palo, J.U., Vainola, R., 2006. The enigma of the landlocked Baikal and Caspian seals addressed through phylogeny of phocine mitochondrial sequences. Biol.J.Linn.Soc. 88, 61-72.

Peigne, S.P., Salesa, M.J., Anton, M., Morales, J., 2005. Ailurid carnivoran mammal Simocyon from the late Miocene of Spain and the systematics of the genus. Acta Palaeontol.Pol. 50, 219-238.

Poe, S., 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Syst.Biol. 52, 423-428.

Posada, D., Crandall, K.A., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Quicke, D.L.J., Jones, O.R., Epstein, D.R., 2007. Correcting the problem of false incongruence due to noise imbalance in the incongruence length difference (ILD) test. Syst.Biol. 56, 496-503.

Rambaut, A., 2008. FigTree. v1.1.2. http://tree.bio.ed.ac.uk/software/figtree/

Roberts, M.S., Gittleman, J.L., 1984. *Ailurus fulgens*. Mamm.Sp. 222, 1-8.

Ruano-Rubio, V., Fares, M.A., 2007. Artifactual phylogenies caused by correlated distributions of substitution rates among sites and lineages: the good, the bad, and the ugly. Syst.Biol. 56, 68-82.

Salesa, M.J., Anton, M., Peigne, S., Morales, J., 2006. Evidence of a false thumb in a fossil carnivore clarifies the evolution of pandas. Proc.Natl.Acad.Sci. 103, 379-382.

Sarich, V.M., Wilson, A.C., 1967. Immunological time scale for hominid evolution. Science 158, 1200-1203.

Sasagawa, I., Takahashi, K., Sakumoto, T., Nagamori, H., Yabe, H., Koabyashi, I., 2003. Discovery of the extinct red panda *Parailurus* (Mammalia, Carnivora) in Japan. J.Vert.Paleo. 23, 895-900.

Sato, J.J., Hosoda, T., Wolsan, M., Suzuki, H., 2004. Molecular phylogeny of Arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of the ferret-badgers and skunks. Zool.Sci. 21, 111-118.

Sato, J.J., Wolsan, M., Suzuki, H., Hosoda, T., Yamaguchi, Y., Hiyama, K., Kobayashi, M., Minami, S., 2006. Evidence from nuclear DNA sequences sheds light on the phylogenetic relationships of pinnipedia: single origin with affinity to musteloidea. Zool.Sci. 23, 125-146.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst.Biol. 51, 492-508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol.Biol.Evol. 16, 1114-1116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246-1247.

Sotnikova, M.V., 2008. A new species of lesser panda *Parailurus* (Mammalia, Carnivora) from the Pliocene of Transbaikalia (Russia) and some aspects of ailurine phylogeny. Paleontological J. 42, 90-99.

Stamatakis, A., 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. Proceedings of IPDPS2006, Rhodos, Greece.

Stamatakis, A., 2006b. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Stanhope, M.J., Czelusniak, J., Si, J.-S., Nickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Binding Protein, with convincing evidence for bat monophyly. Mol.Phylogenet.Evol. 1, 148-160.

Stone, K.D., Flynn, R.W., Cook, J.A., 2002. Post-glacial colonization of northwestern North America by the forest-associated American marten (Martes americana, Mammalia: Carnivora: Mustelidae). Mol.Ecol. 11, 2049-2063.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sunderland, Massachusetts.

Talbot, S.L., Shields, G.F., 1996. A phylogeny of the bears (Ursidae) inferred from complete sequences of three mitochondrial genes. Mol.Phylogenet.Evol. 5, 567-575.

Teeling, E.C., Scally, M., Kao, D.J., Romagnoll, M.L., Springer, M.S., Stanhope, M.J., 2000. Molecular evidence regarding the origin of echolocation and flight in bats. Nature 403, 188-192.

Venta, P.J., Brouillette, J.A., Yuzbasiyan-Gurkan, V., Brewer, G.J., 1996. Gene-specific universal mammalian sequence-tagged sites: application to the canine genome. Biochem.Genet. 34, 321-341.

Wallace, S.C., Wang, X., 2004. Two new carnivores from an unusual late Tertiary forest biota in eastern North America. Nature 431, 556-559.

Wang, X., 1997. New cranial material of *Simocyon* from China, and its implications for phylogenetic relationship to the red panda (*Ailurus*). J.Vert.Paleo. 17, 184-198.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W.J., Townsend, T.M., Reeder, T.W., 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst.Biol. 57, 420-431.

Yang, Z., 1996. Among-site variation and its impact on phylogenetic analyses. TRENDS Ecol.Evol. 11, 367-372.

Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Syst.Biol. 50, 408-424.

Yonezawa, T., Nikaido, M., Kohno, N., Fukumoto, Y., Okada, N., Hasegawa, M., 2007. Molecular phylogenetic study on the origin and evolution of Mustelidae. Gene 396, 1-12.

## Chapter 4

## Novel multi-gene phylogeny of the raccoon family (Procyonidae, Carnivora) reveals extensive morphological convergence[1]

### Introduction

The raccoon family, Procyonidae, is the least studied of the caniform (dog-like) carnivore families. Often represented by two or three species in carnivore phylogenetic studies, to date, no molecular study has focused on the procyonids themselves. The family Procyonidae is sister to Mustelidae (weasels, badgers, otters, and relatives) within the superfamily Musteloidea, which also includes the families Ailuridae (monotypic red panda, *Ailurus fulgens*) and Mephitidae (skunks). The first complete phylogeny of the recent procyonids (Decker and Wozencraft, 1991) based on 129 morphological characters delineated two subfamilies, Potosinae and Procyoninae, which were upheld (as the tribal designations Potosini and Procyonini) in the morphological phylogeny of Baskin (2004), including both extant and fossil procyonids. The red panda has sometimes been included in the Procyonidae as a third subfamily (Simpson, 1945; Baskin, 1998, 2003), but while the phylogenetic affinity of the red panda within the musteloids is unknown (Flynn et al., 2000; Delisle and Strobeck, 2005; Flynn et al., 2005; Flynn and Wesley-Hunt, 2005; Fulton and Strobeck, 2006; Sato et al., 2006), it is no longer generally accepted as a member of the Procyonidae. The subfamily Potosinae contains two genera, *Potos* (kinkajou) and *Bassaricyon* (olingos), while Procyoninae contains three: *Bassariscus* (ringtails), *Nasua* (coatis), and *Procyon* (raccoons). A putative sixth genus contains the monotypic mountain coati, *Nasuella olivaceae*. Found only in the Andes, *N. olivaceae* is smaller than, but otherwise similar to, the ring-tailed coati, *Nasua nasua*, and is often considered a species of *Nasua*, rather than a distinct genus (Nowak, 1991).

The Procyonidae have one of the poorest fossil records of any of the carnivoran families (Baskin, 1982), exemplified by the fact that no fossils resembling *Bassaricyon (Bassaricyonoides)* or *Potos (Parapotos)* were described until recently (Baskin, 2003). The earliest procyonid, *Pseudobassaris riggsi*, is known from the late Oligocene in France (Wolsan, 1993; Wolsan and Lange-Badre, 1996) approximately 28 million years ago (see Sato et al., 2003 and references therein); New World procyonids first occur in early Miocene North America, and

[1] A version of this paper has been published. Fulton, T.L. and C. Strobeck. 2007. Molecular Phylogenetics and Evolution 43, 1171-1177.

late Miocene South America, representing the first South American carnivorans (Baskin, 1998, 2003, 2004; Flynn and Wesley-Hunt, 2005). Divergences leading to extant genera are thought to have occurred at approximately this time in the New World, from early Miocene to early Pliocene (Simpson, 1945; Baskin, 1989, 2003, 2004). However, relationships among fossil procyonids were relatively unclear until recently, as phylogenetic studies of these taxa have had some problematic limitations. Baskin (1982) provided the first phylogenetic study of the Procyonidae (including fossil taxa), but included *Bassariscus* as an outgroup and did not include *Potos*. This study was also heavily biased toward dental characters (~80%), an unavoidable problem as many fossil genera and species are known only from dentition. Outgroup selection was corrected in Baskin (1989), but *Bassaricyon* was then omitted. Character polarity may also be problematic when Ailuridae is included as a subfamily (Ailurinae) within Procyonidae (with recent Procyonidae as Procyoninae), to the exclusion of Mustelidae (Baskin, 1998), given more recent contradictory molecular (Flynn et al., 2000; Delisle and Strobeck, 2005; Flynn et al., 2005; Fulton and Strobeck, 2006; Sato et al., 2006) and fossil evidence (Wolsan, 1993; Salesa et al., 2006). Sampling and outgroup selection issues are corrected in the most recent procyonid phylogeny (Baskin, 2004) based on 40 morphological characters (75% dental characters), for all extant (except *Nasuella*) and extinct New World genera, including the recently described 'potosin' fossils (Baskin, 2003). Though unavoidably heavily based on dental characters (see above discussion), a single most parsimonious tree was obtained, supporting the distinction between Potosinae and Procyoninae, as well as a sister relationship of New World taxa with the early Miocene European *Broiliana* (Baskin, 2004).

Here, I present the first molecular-based phylogeny investigating the relationships within the extant Procyonidae using three nuclear and three mitochondrial (mt) genes.

## Materials and methods

### *Samples, amplification, and sequencing strategy*

17 samples representing 7 procyonids and 10 additional caniform carnivores as outgroup taxa were included in this study (Table 4-1). Sequences were obtained for two sequence-tagged sites (STS), *cholinergic receptor, nicotinic, alpha polypeptide 1 precursor (CHRNA1)* intron 8, and *growth hormone receptor (GHR)* intron 9; one nuclear exon, *interphotoreceptor retinoid-binding protein (IRBP)* exon 1; and three mitochondrial genes, *cytochrome c oxidase I (COI),*

*NADH dehydrogenase subunit 2 (ND2), and cytochrome b (CYTB).* All sequences for *Bassaricyon gabbii* (except *CYTB*), *COI* and *ND2* for *Bassariscus astutus* and *Potos flavus*, and *CYTB, COI,* and *ND2* for *Nasua narica, Enhydra lutris,* and *Mustela nivalis* were newly generated for this study. Amplification PCR procedures were used for *CHRNA1* and *GHR* as in Koepfli and Wayne (2003), for *IRBP* as in Stanhope et al. (1992), and for *ND2, CYTB,* and *COI* as in Delisle and Strobeck (2002). Bi-directional sequencing was performed using BigDye v.3.1.1 (Applied Biosystems) following the manufacturer's protocol using the amplification primers. Fragments were resolved using an Applied Biosystems 3730 sequencer. All other sequences were obtained from GenBank (see Table 4-2 for accession numbers) and new sequences have been entered into GenBank (accessions DQ533934 to DQ533952).

*Phylogenetic analyses*

Sequences were analyzed, basecalled, and aligned using Foundation Data Collection v.3.0 and Sequence Navigator v.1.0.1 (both from Applied Biosystems). Heterozygous sites in nuclear regions (equal peak heights in electropherograms observed in both directions of sequence) were coded as polymorphisms. The new *B. gabbii* sequences were added by eye to existing alignments (Fulton and Strobeck, 2006) available in TreeBASE (accession S1532) for the nuclear genes and mt sequences were aligned by eye. Both STS regions contained informative insertion/deletion events (indels); these were coded as 0/1 (absence/presence of DNA) regardless of length (Barriel, 1994) and included in maximum parsimony analyses. To ensure congruence among data sets prior to concatenation, partition homogeneity tests (ILD test, Farris et al., 1995) were performed in PAUP* v.4.10b (Swofford, 2003) comparing each gene to the other five genes combined, pair-wise comparisons between all genes, and between coding positions for *CYTB, ND2, COI,* and *IRBP*. All tests were non-significant at $\alpha$=0.05. Due to limitations of the ILD test to accurately assess congruence and therefore, data combinability, (Dolphin et al., 2000; Barker and Lutzoni, 2002; Darlu and Lecointre, 2002), maximum parsimony bootstrap analysis of each gene partition separately was also performed, using 1000 pseudoreplicates with 10 random sequence addition replicates. No significant (≥70% BP) conflicts were observed, thus, gene partitions were combined where appropriate. ModelTest 3.7 (Posada and Crandall, 1998) was employed for each gene separately and for all regions concatenated to determine the most appropriate DNA evolution model, as selected by hierarchical likelihood ratio tests.

Maximum parsimony heuristic searching with TBR branch-swapping on 10 random sequence addition replicate starting trees was employed in PAUP* (Swofford, 2003), including indels and delineating alignment gaps as 'missing'. Maximum parsimony bootstrap (MP BP) heuristic searching was also employed in PAUP*, using the same search parameters for 10 000 bootstrap replicates. An iterative approach to maximum likelihood (ML) heuristic searching (Sullivan et al., 2005 and references therein) was employed in PAUP*, whereby the parameters estimated from ModelTest on the MP topology were fixed and the topology was estimated using heuristic searching (TBR branch swapping) from a randomly generated topology. Then, holding the topology constant, parameters were estimated. This iteration was repeated twice to ensure that the likelihood score and parameter estimates had stabilized. Maximum likelihood bootstrapping (ML BP) was also performed, holding all model parameters constant as estimated on the final ML tree (Douady et al., 2003), using neighbour-joining starting trees and TBR branch swapping for 100 pseudoreplicates. MtDNA and nuclear partitions were analyzed in the same way, but with 50 pseudoreplicates. Bayesian analysis was performed using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), with each region as its own separate, unlinked, partition with parameters free to vary under the appropriate DNA evolution model as selected using ModelTest. Four chains (using default temperature = 0.2 for three heated chains) were run for 2 000 000 generations, sampling the cold chain every 100 generations. The first 10% of the samples were discarded as burn-in after visualization using the sump command in MrBayes. Two runs were performed to ensure proper convergence of each chain – average standard deviation of split frequencies after two million generations was ~0.0009 and the potential scale reduction factor was approximately 1.0 for all parameters.

**Results**

*General sequence results and insertion/deletion information*

The complete alignment of the six regions yielded a total of 5918 aligned base pairs (bp), consisting of *IRBP* (1187bp), *COI* (1545bp), *CYTB* (1140bp), *ND2* (1044), *GHR* (632bp), and *CHRNA1* (370bp). *GHR* contained five parsimony-informative indels; *CHRNA1* contained two (Figure 4-1a). All indels supported clades that were strongly supported by all other methods, except a homoplasious 4 bp deletion in *GHR* shared by *Potos flavus* and *Lobodon carcinophagus*. Length variation was observed in two of the newly generated sequences. *COI* is two amino acids

(a.a.) longer in *Potos flavus* and *CYTB* is two a.a. longer in *Nasua narica*. In *P. flavus*, the TAA termination found in all other taxa examined for *COI* is replaced by CAA (Gln), then GAA (Glu), and an AGG termination (vertebrate mitochondrial code). For *CYTB* in *N. narica*, the AGA termination observed in all other taxa (except *Bassaricyon gabbii*, which has GGG, (Gly)) is replaced by AAA (Lys), then ACT (Thr), and a TAA termination. *CYTB* sequence for *B. gabbii* was not generated by this study, so the length of this gene cannot be determined. Both sequences showing length variation were sequenced multiple times from both directions for confirmation, but additional bases were not included in analyses.

All tree search methods using all data yielded the same topology (Figure 4-1) with the exception of the placement of the red panda under the parsimony criterion, as described below. The MP tree required 7186 steps, with consistency index (CI) = 0.4887 and retention index (RI) = 0.3536. Excluding all mtDNA third codon positions, MP analysis recovered the same topology but improved the CI to 0.6118 and RI to 0.4903. Including only nuclear genes yielded CI and RI values of ~0.75, illustrating their lower level of homoplasy and consequent utility in phylogenetic reconstruction. ML estimation was performed under the general time reversible (GTR) model with estimation of the proportion of invariant sites (I) and gamma distributed rate variation (Γ). Ln likelihood was -36759.90893, estimated base frequencies were A=0.291, C=0.280, G=0.179, T=0.250, I=0.436499, and gamma shape = 0.723478. Transitions and transversions were observed at drastically different rates within the concatenated data set: based on the reference GT=1.0, transversion rates ranged from CG=0.68 to AC=2.35, while transition rates were AG=8.87 and CT=20.06. On average, Ti:Tv was approximately 9:1.

*Phylogeny of the Procyonidae*

All analysis methods strongly support three lineages within Procyonidae: *Potos* as most basal, a *Nasua+Bassaricyon* clade, and a *Procyon+Bassariscus* clade (Figure 4-1a). All nodes within the family were supported by BPP of 1.0 and 100% MP and ML BP support, except the clade joining the four procyonid genera excluding *Potos*, which was still strongly supported by MP BP=96, ML BP=100, BPP=1.0, and a four bp deletion in *GHR*. The *Bassaricyon + Nasua* clade was supported by a 12 bp deletion in *CHRNA1*.

*Alternate topologies recovered for familial relations within Musteloidea*

Likelihood-based analyses (ML and Bayesian) and MP analysis recovered different, well-supported placements of the red panda (Ailuridae) within Musteloidea (Figures 4-1b and 4-1c). Likelihood-based methods weakly supported (BPP=0.81, ML BP=43) the branching order Ailuridae, Mephitidae, then Procyonidae + Mustelidae, whereas parsimony analysis strongly supported (MP BP=99) Ailuridae and Mephitidae as sister. Parsimony also recovered the same topology when only nuclear DNA was analyzed, indicating that mtDNA signal supporting a sister relationship between Mephitidae and Ailuridae (results not shown) is not 'swamping' the parsimony analysis due to the inclusion of approximately 5-fold more parsimony-informative sites. ML analysis of only nuclear DNA supports (ML BP=61) Mephitidae as the most basal musteloid, in contrast to only mtDNA, which very weakly (ML BP=33) recovers the same topology as the concatenated analysis. Tests for base composition differences ($\chi^2$ for homogeneity of base frequencies in PAUP*) including all taxa with sequence either concatenated or partitioned by nuclear or mtDNA indicate significant base composition bias (all $p<0.001$), although analyses including each gene partition separately do not (all $p>0.05$). *Ailurus fulgens* and *M. mephitis* do not have significantly different base composition from one another (all sequence p=0.78, mt p=0.73, nuclear p=0.91). All comparisons between *A. fulgens* and/or *M. mephitis* with other musteloids (separately or in nearly all combinations) using the nuclear DNA partition were not significant (all $p>0.95$), while comparisons of mtDNA were, or were close to, significance (all $p\sim0.05$).

**Discussion**

*Phylogeny of the Procyonidae*

All analysis methods yielded a single, strongly supported topology for the Procyonidae (Figure 4-1), comprised of three lineages: the kinkajou (*Potos flavus*), olingos (*Bassaricyon*) + coatis (*Nasua*), and ringtails (*Bassariscus*) + raccoons (*Procyon*), with the latter two clades as sister. All nodes within the Procyonidae were supported by 100% MP and ML BP support and BPP=1.0, except the clade including all genera but *Potos*, which was still strongly supported by MP BP=96, ML BP=100, BPP=1.0 and a 4 base pair deletion in *GHR*. A 12 base pair deletion in *CHRNA1* was synapomorphic for the *Bassaricyon* + *Nasua* clade, providing strong additional

support for the recovered topology. However, this phylogeny for the recent procyonids is highly discordant with the present subfamilial designations of Potosinae (*Potos, Bassaricyon*) and Procyoninae (*Procyon, Nasua, Bassariscus*) as supported by morphology (Decker and Wozencraft, 1991; Baskin, 2004). The morphological characters used to define these clades may, in fact, be resultant from convergent evolution as opposed to shared evolutionary history.

Habitat and diet are highly variable across the Procyonidae. *Potos* and *Bassaricyon* both contain arboreal, mainly frugivorous, nocturnal species that inhabit the same niche – the upper canopy of the Central American rainforests – to the extent that they are often competitors for resources (Kays, 2000). *Potos* are the most autapomorphic of all the procyonids and are highly adapted for their arboreal, frugivorous, li*FE*Style with a fully prehensile tail and elongate tongue for obtaining nectar and honey (Ford and Hoffman, 1988; Nowak, 1991). *Bassariscus* are omnivorous and primarily arboreal, with *B. astutus* being the only caniform carnivore to have semi-retractile claws (Poglayen-Neuwall and Toweill, 1988). *Procyon* and *Nasua* are considerably more terrestrial than the other genera. Like *Bassariscus*, *Procyon* species are omnivorous, while *Nasua* are insectivorous. Several characters related to shared ecology were used to define the two morphology-based subfamilies, but in light of our molecular evidence, it is possible that some of these characters are homoplasious, rather than synapomorphic. *Bassaricyon* and *Potos* were placed as sister based on a reduced number of cusps on molars, four auditory-related characters including short, rounded (vs. long, pointed) ears, three cranial characteristics (two of which are orbit-related), and a sharply-angled (vs. straight) acromion process in the shoulder. Of the nine characters used to define the Procyoninae by Decker and Wozencraft (1991), the three molar cusp characters may be problematic (Baskin, 2004) and a fourth character, the presence of banded tail rings, is more consistent with our topology since although less prominent, are also present in some *Bassaricyon* species. Some of the characters used to group *Bassaricyon* and *Potos* have been shown to evolve in parallel in other species, such as reduction of cusps on molars and grooved canines in frugivores and the forward placement of the eyes for increased stereoscopic vision in arboreal species (Decker and Wozencraft, 1991). For example, these 'frugivorous' dental characteristics are also observed in the small-toothed palm civet, *Arctogalidia trivirgata,* a distantly related feliform carnivore (Decker and Wozencraft, 1991). *Potos* has also been described as "convergently resembling" another viverrid, *Arctictis binturong*, due to its arboreal and nocturnal adapations including a prehensile tail (Ford and Hoffman, 1988). Because morphological studies must rely heavily on dental and cranial characteristics, they can be strongly affected by undetected parallel evolution interpreted as shared evolution, as is likely the

case of the Procyonidae. It is also possible, though less likely, that some of the characteristics shared by *Potos* and *Bassaricyon* are actually plesiomorphic, as most fossil procyonids were hypocarnivorous (omnivorous or frugivorous) and potentially arboreal (Baskin, 2003).

*Alternate phylogenetic placements of the red panda*

All analyses yielded the same topology (Figure 4-1), with the exception of the placement of the 'enigmatic' red panda, *Ailurus fulgens*. Ailuridae has been proposed as sister to Mephitidae based primarily on mtDNA sequence analysis (Flynn et al., 2000; Delisle and Strobeck, 2005), as sister to Procyonidae+Mustelidae based on analysis of nuclear DNA alone (Fulton and Strobeck, 2006; Sato et al., 2006), or as the most basal musteloid lineage when both types of DNA markers are combined (Flynn et al., 2005). Consistent with the last hypothesis, our model-based analyses of combined mt and nuclear DNA weakly support (BPP=0.81, ML BP=43) a sister relationship between Mephitidae and Procyonidae+Mustelidae. This is in contrast to my parsimony analysis, which recovered Ailuridae and Mephitidae as sister with unusually strong support (MP BP=99). When nuclear and mtDNA partitions were analyzed separately, MP always recovered Ailuridae+Mephitidae, but ML differed based on the partition analyzed. ML analysis of the three nuclear genes supported (ML BP=61) Mephitidae as the most basal musteloid lineage, consistent with previous 'nuclear-only' studies that rejected both other hypotheses (SH test, Fulton and Strobeck, 2006). Conversely, ML analysis of the three mt genes very weakly supported (ML BP=33) Ailuridae as the most basal lineage, as in the combined analysis. The incongruent topologies from parsimony and likelihood mtDNA analyses are the same as those of Flynn et al.'s (2000) analysis of three mt and one nuclear gene. Given that my concatenated data ML analysis is consistent with the mtDNA ML analysis, it is possible that while support is increased in the concatenated analysis, the mtDNA signal is overwhelming the combined analysis, biasing the species level interpretations toward what could be considered a single gene tree and leading to the inconsistency with the 'nuclear' topology.

Differences between parsimony and model-based analyses are potentially attributed to some problems to which parsimony can be more susceptible. Divergence between Ailuridae and Mephitidae occurred early in the radiation of the musteloid families ~30 MYA (Salesa et al., 2006), over a short period of evolutionary time, yielding relatively long, unbroken branches (Figure 4-1a). These long branches in combination with the very short internal branch joining the two create a 'Felsenstein zone' type topology, for which parsimony has been shown to be

inconsistent, increasingly supporting an incorrect tree (see Felsenstein, 2004). This problem may be exacerbated in this study due to the inclusion of only a single mephitid, as studies including more skunks recover lower parsimony based support (3 mephitids: Fulton and Strobeck, 2006) or remain unresolved (4 mephitids: Flynn et al., 2005) for this relationship, likely due to the additional mephitids 'breaking up' one of the long branches. Though Ailuridae+Mephitidae is obtained via parsimony regardless of codon partitioning strategy (Chapter 3), it remains possible that transition mutational saturation may also be having an effect, supported by the lower consistency index of mtDNA compared to nuclear DNA alone and the propensity for mt genes to more quickly accumulate mutations, especially at the third codon position. Base composition bias may also lead to erroneous groupings, whereby taxa with similar base composition may be grouped together (Lockhart et al., 1994). Differences in base composition were observed in the mtDNA partition within Musteloidea, but not between *Ailurus* and *Mephitis*, indicating that this may be contributing to their placement as sister. Some, or all, of these factors may be contributing to the strong parsimony bootstrap support seen for an Ailuridae+Mephitidae clade. Further study into the conflict between tree estimation methods and potentially differing mt and nuclear DNA signal and is required. Combining increased molecular data with morphological information from both extant and extinct taxa will be necessary to more confidently address the phylogenetic affinity of the red panda.

**Conclusions**

I provide a novel phylogeny for the recent Procyonidae, and, given the strong support and congruence between genes and analysis methods, suggest that the present morphology-based subfamilial designations of Potosinae and Procyoninae are inappropriate. High variation in habitat and diet across procyonids appears to have heavily impacted the evolution of species, making the Procyonidae an excellent example of convergent evolution. I suggest a re-evaluation of the morphology of both extant and fossil taxa, in hopes of reconciling or better understanding the discrepancy between the molecular and morphological topologies.

**Table 4-1.** Species included in study.

| Classification | | Species Name | Common Name |
|---|---|---|---|
| Caniformia | | | |
| | Canidae | *Canis lupus* | Gray wolf |
| | Arctoidea | | |
| | Ursidae | *Ursus maritimus* | Polar bear |
| | Musteloidea | | |
| | Procyonidae | *Potos flavus* | Kinkajou |
| | | *Procyon lotor* | North American raccoon |
| | | *Procyon cancrivorous* | Crab-eating raccoon |
| | | *Nasua nasua* | Ring-tailed coati |
| | | *Nasua narica* | White-nosed coati |
| | | *Bassariscus astutus* | Ringtail |
| | | *Bassaricyon gabbii* | Olingo |
| | Mustelidae | | |
| | Lutrinae | *Enhydra lutris* | Sea otter |
| | Taxidiinae | *Taxidea taxus* | American badger |
| | Mustelinae | *Martes americana* | American marten |
| | | *Mustela nivalis* | Least weasel |
| | Mephitidae | *Mephitis mephitis* | Striped skunk |
| | Ailuridae | *Ailurus fulgens* | Red panda |
| | Pinnipedia | | |
| | Odobenidae | *Odobenus rosmarus* | Walrus |
| | Phocidae | *Lobodon carcinophagus* | Crabeater seal |

**Table 4-2.** GenBank accession numbers for species included in study. Blank cells indicate that sequence was unavailable due to absence of sequence in GenBank and no DNA sample available. References for sequences obtained from GenBank are noted by: [a]Yu et al. (2004) [b]Sato et al. (2003) [c]Sato et al. (2004) [d]Fulton and Strobeck (accepted) [e]Koepfli and Wayne (2003) [f]Davis et al. (2004) [g]Delisle and Strobeck (2005) [h]Delisle and Strobeck (2002) [i]Zhang and Ryder (1993) [j]Ledje and Arnason (1996) [k]Koepfli and Wayne (1998) [m]Stone et al. (2002). All other sequences were generated by this study.

| Species | *IRBP* | *GHR* | *CHRNA1* | *ND2* | *CYT B* | *COI* |
|---|---|---|---|---|---|---|
| *Ursus maritimus* | AY303843[a] | DQ205798[d] | DQ205725[d] | NC_003428[h] | NC_003428[h] | NC_003428[h] |
| *Procyon lotor* | AB082981[b] | AF498207[e] | AF498152[e] | AY598590[g] | X94930[a] | AY598575[g] |
| *Procyon cancrivous* | AB109332[c] | | | | | |
| *Nasua narica* | DQ205878[d] | DQ205802[d] | DQ205729[d] | DQ533944 | DQ533940 | DQ533934 |
| *Nasua nasua* | AY525031[a] | | | | | |
| *Bassariscus astutus* | DQ205879[d] | AF498206[e] | DQ205730[d] | DQ533945 | AF498159[e] | DQ533935 |
| *Potos flavus* | DQ205880[d] | DQ205803[d] | DQ205731[d] | DQ533946 | L21876[i] | DQ533936 |
| *Bassaricyon gabbii* | DQ533952 | DQ533951 | DQ533950 | DQ533947 | X94931[j] | DQ533937 |
| *Ailurus fulgens* | DQ205881[d] | DQ205804[d] | DQ205732[d] | AY598524[g] | X94919[j] | AY598520[g] |
| *Mephitis mephitis* | AB109331[c] | DQ205805[d] | DQ205733[d] | AY598535[g] | X94927[j] | AY598531[g] |
| *Enhydra lutris* | AB082978[b] | AF498186[e] | AF498131[e] | DQ533948 | DQ533942 | DQ533938 |
| *Taxidea taxus* | DQ205885[d] | AF498203[e] | AF498148[e] | AY598568[g] | AF057132[k] | AY598564[g] |
| *Martes americana* | AB082963[b] | AF498196[e] | AF498141[e] | AY598546[g] | AY121352[m] | AY598542[g] |
| *Mustela nivalis* | AB082973[b] | DQ205809[d] | DQ205737[d] | DQ533949 | DQ533943 | DQ533939 |
| *Odobenus rosmarus* | DQ205892[d] | DQ205815[d] | DQ205741[d] | AY377281[f] | X82299[h] | AY377148[f] |
| *Lobodon carcinophagus* | DQ205896[d] | DQ205819[d] | DQ205745[d] | AY377268[f] | AY377321[f] | AY377130[f] |
| *Canis lupus* | DQ205907[d] | DQ205836[d] | DQ205757[d] | AY598501[g] | AY598499[g] | AY598496[g] |

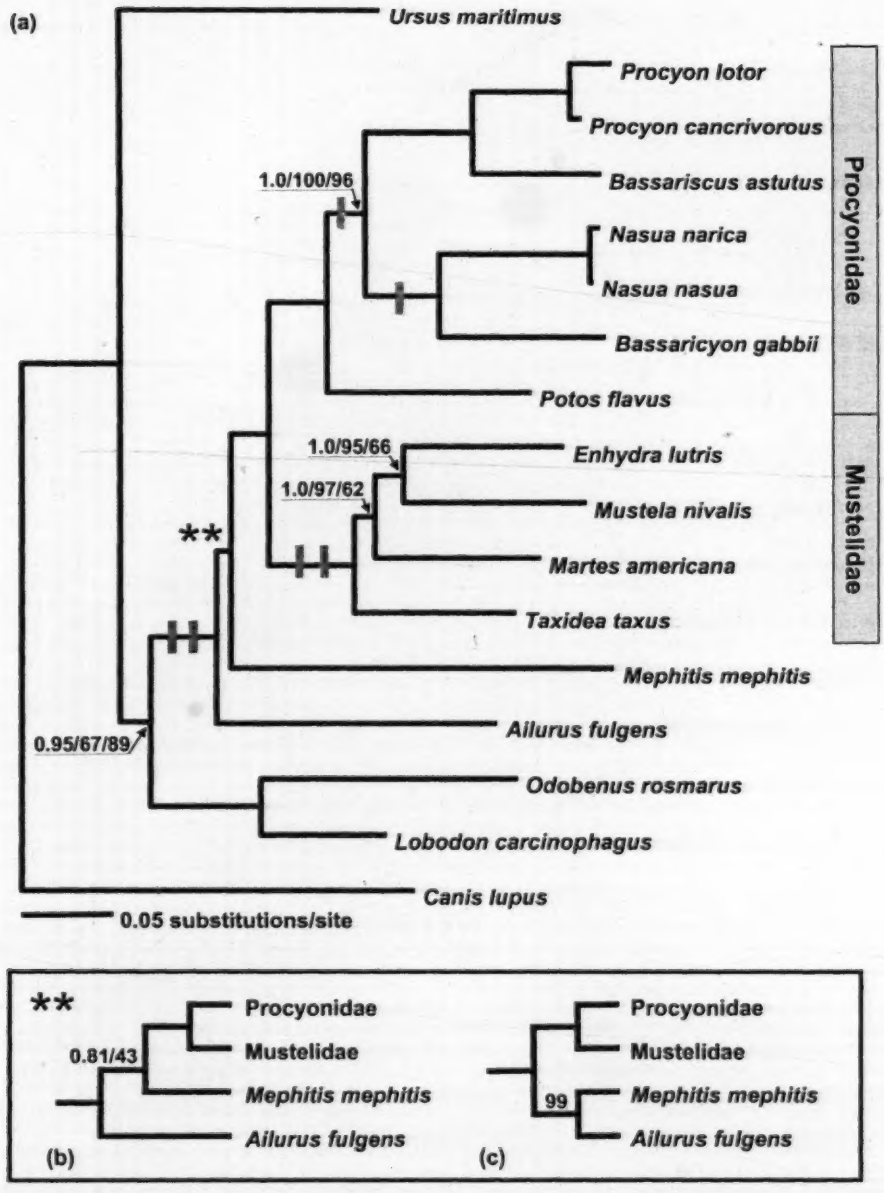


**Figure 4-1.** Phylogeny of the Procyonidae. (a) Maximum likelihood topology of combined nuclear and mitochondrial genes, including optimized branch lengths, also recovered by Bayesian analysis and, with a single marked exception (**), by maximum parsimony. All nodes were supported by Bayesian posterior probability (BPP) of 1.0, maximum likelihood bootstrap (ML BP) of 100%, and maximum parsimony bootstrap (MP BP) of 100% except where noted as indicated by BPP / ML BP / MP BP. Grey vertical bars mark the position that indels are mapped onto the tree. (b) The placement of *Ailurus fulgens* and *Mephitis mephitis* as recovered by ML and Bayesian analysis, including BPP / ML BP support. (c) The placement of *A. fulgens* and *M. mephitis* as recovered by parsimony, including MP BP support.

**References**


Barker, F.K., Lutzoni, F.M., 2002. The utility of the incongruence length difference test. Syst.Biol. 51, 625-637.

Barriel, V., 1994. Molecular Phylogenies and How to Code Insertion/Deletion Events. Life Sci. 317, 693-701.

Baskin, J.A., 1982. Tertiary Procyoninae (Mammalia: Carnivora) of North America. J.Vert.Paleo. 2, 71-93.

Baskin, J.A., 1989. Comments on New world Tertiary Procyonidae (Mammalia: Carnivora). J.Vert.Paleo. 9, 110-117.

Baskin, J.A., 1998. Procyonidae. In: Janis, C.M., Scott, K.M., Jacobs, L.J. (Eds.), Evolution of Tertiary mammals of North America. Volume 1: Terrestrial carnivores, ungulates, and ungulatelike mammals. Cambridge University Press, New York, NY, pp. 144-151.

Baskin, J.A., 2003. New procyonines from the Hemingfordian and Barstovian of the Gulf coast and Nevada, including the first fossil record of the Potosini (Chapter 6). In: Flynn, L.J. (Ed.), Vertebrate fossils and their context: contributions in honor of Richard H. Tedford. Bull.Am.Mus.Nat.Hist., Number 279, New York, NY, pp. 125-146.

Baskin, J.A., 2004. Bassariscus and Probassariscus (Mammalia, Carnivora, Procyonidae) from the early Barstovian (middle Miocene). J.Vert.Paleo. 24, 709-720.

Darlu, P., Lecointre, G., 2002. When does the incongruence length difference test fail? Mol.Biol.Evol. 19, 432-437.

Decker, D.M., Wozencraft, W.C., 1991. Phylogenetic analysis of recent procyonid genera. J.Mamm. 72, 42-55.

Delisle, I., Strobeck, C., 2002. Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. Mol.Biol.Evol. 119, 357-361.

Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Mol.Phylogenet.Evol. 37, 192-201.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. Mol.Phylogenet.Evol. 17, 401-406.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol.Biol.Evol. 20, 248-254.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Constructing a significance test for incongruence. Syst.Biol. 44, 570-572.

Felsenstein, J., 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland, Massachusetts.

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M.A., 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst.Biol. 54, 317-337.

Flynn, J.J., Nedbal, M.A., Dragoo, J.W., Honeycutt, R.L., 2000. Whence the red panda? Mol.Phylogenet.Evol. 17, 190-199.

Flynn, J.J., Wesley-Hunt, G.D., 2005. Carnivora. In: Rose, K.D., Archibald, J.D. (Eds.), The rise of placental mammals. The John Hopkins University Press, Baltimore, Maryland, pp. 175-198.

Ford, L.S., Hoffman, R.S., 1988. Potos flavus. Mamm.Species 321, 1-9.

Fulton, T.L., Strobeck, C., 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. Mol.Phylogenet.Evol. 41, 165-166-181.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Kays, R.W., 2000. The behavior and ecology of olingos (Bassaricyon gabbii) and their competition with kinkajous (Potos flavus) in central Panama. Mammalia 64, 1-9.

Koepfli, K.-P., Wayne, R.K., 2003. Type I STS markers are more informative than cytochrome b in phylogenetic reconstruction of the Mustelidae (Mammalia: Carnivora). Syst.Biol. 52, 571-593.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering Evolutionary Trees under a More Realistic Model of Sequence Evolution. Mol.Biol.Evol. 11, 605-612.

Nowak, R.M., 1991. Walker's Mammals of the World. Fifth Edition. The John Hopkins University Press, Baltimore.

Poglayen-Neuwall, I., Toweill, D.E., 1988. Bassariscus astutus. Mamm.Species 327, 1-8.

Posada, D., Crandall, K.A., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Salesa, M.J., Anton, M., Peigne, S., Morales, J., 2006. Evidence of a false thumb in a fossil carnivore clarifies the evolution of pandas. Proc.Nat.Acad.Sci. 103, 379-382.

Sato, J.J., Hosoda, T., Wolsan, M., Tsuchiya, K., Yamamoto, M., Suzuki, H., 2003. Phylogenetic relationships and divergence times among Mustelids (Mammalia: Carnivora) based on nucleotide sequences of the Nuclear Interphotoreceptor Retinoid Binding Protein and mitochondrial cytochrome b genes. Zool.Sci. 20, 243-264.

Sato, J.J., Wolsan, M., Suzuki, H., Hosoda, T., Yamaguchi, Y., Hiyama, K., Kobayashi, M., Minami, S., 2006. Evidence from nuclear DNA sequences sheds light on the phylogenetic relationships of pinnipedia: Single origin with affinity to musteloidea. Zool.Sci. 23, 125-146.

Simpson, G.G., 1945. The principles of classification and a classification of mammals. Bull.Am.Mus.Nat.Hist. 85, 1-350.

Stanhope, M.J., Czelusniak, J., Si, J.S., Nickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Binding Protein, with convincing evidence for bat monophyly. Mol.Phylogenet.Evol. 1, 148-160.

Sullivan, J., Abdo, Z., Joyce, P., Swofford, D.L., 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. Mol.Biol.Evol. 22, 1386-1392.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). version 4. Sinauer Associates, Sunderland, Massachusetts.

Wolsan, M., 1993. Phylogeny and classification of early European Mustelida (Mammalia, Carnivora). Acta Theriol. 38, 345-384.

Wolsan, M., Lange-Badre, B., 1996. An arctomorph carnivoran skull from the Phosphorites du Quercy and the origin of procyonids. Acta Palaeontol.Polon. 41, 277-298.

## Chapter 5

## Multiple individuals, multiple nuclear loci, and the importance of partitioning mitogenomic datasets in determining the phylogeny of true seals (Phocidae, Pinnipedia)

### Introduction

Nearly two hundred years after being defined as a group (Illiger, 1811), the phylogeny of the Pinnipedia remains to be completely resolved. But with the increasing ease of DNA sequencing, it has become feasible to obtain sequence information from numerous independent loci to address pinniped phylogeny. As uncertainties remain about relationships at every taxonomic level from superfamily to species, DNA markers must be selected to encompass a range of evolutionary rates and inheritance patterns, providing information to resolve both the deepest and most shallow clades. While increasing the heterogeneity and size of a dataset can improve phylogenetic estimation, it also increases the difficulty in modeling such heterogeneity (Delsuc et al., 2005; Nishihara et al., 2007; Rodriguez-Ezpeleta et al., 2007).

Unlinked nuclear genes represent independent evolutionary histories, and while mitochondrial genes do not represent such independent evolutionary units, different genes, codon positions, and even nucleotide sites are likely not evolving in a similar enough manner to be adequately explained by a single evolutionary model and associated set of parameters. There are three widely applied methods of accounting for evolutionary heterogeneity in the mitochondrial genome. The first is simply to identify sections of the data that violate the assumptions of the applied evolutionary model and exclude them from analysis. The third codon position nucleotides often exhibit differing base composition across taxa. Violating model assumptions of base composition stationarity across taxa can lead to artifactual taxon groupings based on similar base composition, as opposed to true shared evolutionary history (Jermiin et al., 2004). A second, related, method is to mask the problem. Re-coding DNA sequences as either purines or pyrimidines (RY coding) to eliminate transitions (and emphasize transversions) or translating the DNA sequence into its inferred amino acid sequence can sometimes compensate for base composition bias across taxa. However, amino acid sequence is still not immune to problems of compositional bias (Gibson et al., 2005). The primary use of recoding is to avoid saturation effects that are also difficult to model, as saturation results in a loss of phylogenetic information and an increase in dataset homoplasy. Both RY and amino acid coding function to reduce noise in the dataset, but in doing so, information about recent divergences provided by synonymous

mutations is lost. The third method for accounting for data heterogeneity is to subdivide, or partition, the data into similarly evolving units. Traditionally, mitochondrial loci are partitioned by gene, by codon position, or both. A recent method of partitioning DNA sequence is to use consistency index as a proxy for rate (Kjer and Honeycutt, 2007). In this method, hereon in referred to as 'partitioning by rate', characters are grouped together based on their maximal consistency index from a number of bootstrap replicates. Characters with the lowest consistency index (CI) represent the 'fastest' sites and those with the highest CI represent the 'slowest'. Identifying a dataset that minimizes model assumption violations and selecting a partitioning strategy that best defines similarly evolving units is an important component of phylogenetic reconstruction and can potentially lead to a reduction of phylogenetic artifacts (Nishihara et al., 2007).

How does one choose the 'best' method of partitioning and combining data? Although this is not a new question (Bull et al., 1993), it remains an important problem in phylogenetics. In this study, various mitochondrial partitioning strategies are compared using Bayes factors. When two different models (in this case, partitioning strategies) are applied to the same dataset, the ratio of the marginal likelihoods (the Bayes factor) can be interpreted as the relative ability of each model to predict the data (Kass and Raftery, 1995). The 'best' method of partitioning the mitochondrial data is compared both with excluding sections of the dataset that violate model assumptions and with amino acid analyses. In turn, the mitochondrial results are compared to and combined with the nuclear results and data and all molecular results are considered in the context of previous morphological hypotheses.

In addition to identifying optimal methods of phylogenetic reconstruction for this dataset, a second approach is taken towards resolving rapid radiations and addressing taxonomic issues. In the vein of combining a coalescent approach with phylogenetic reconstruction, two individuals per species are included, except in four cases where multiple samples were unavailable. For recent rapid radiations, including multiple individuals may allow identification of coalescence events within a species, allowing better differentiation between inter- and intraspecific variation. Within the northern seal subtribe Phocina, rapid species radiations combined with the large, somewhat unstructured populations of *Pusa hispida* (King, 1983; Palo et al., 2001; Sasaki et al., 2003), mean that incomplete lineage sorting may be problematic for resolution. Increasing both the number of loci and the number of individuals should help to disentangle these close relationships (Maddison and Knowles, 2006; Carstens and Knowles, 2007). Using multiple individuals for all phocid species also sets a guideline for inter- and intra-

species and genus genetic differentiation to address taxonomy from a genetic perspective.

Three phylogenetic questions are addressed. First, the higher-level taxonomy (Table 5-1) is investigated, from pinniped superfamily relationships to subfamily and tribal relationships within Phocidae (true, or earless, seals). There are three pinniped families: Phocidae, Odobenidae (monotypic walrus), and Otariidae (fur seals and sea lions). Though both morphological and molecular evidence place the latter two families together in the superfamily Otarioidea (Repenning et al., 1979; King, 1983; Ledje and Arnason, 1996; Dragoo and Honeycutt, 1997; Flynn and Nedbal, 1998; Davis et al., 2004; Delisle and Strobeck, 2005; Flynn et al., 2005; Arnason et al., 2006; Fulton and Strobeck, 2006; Higdon et al., 2007), a closer relationship of Odobenidae to Phocidae (together called Phocamorpha) has also been proposed (Wyss, 1987, 1988; Wyss and Flynn, 1993; Berta and Wyss, 1994). There are two subfamilies within Phocidae (King, 1966), Phocinae (northern seals) and Monachinae (Antarctic, monk, and elephant seals), and each subfamily is divided into three tribes (Table 5-1, *sensu* Burns and Fay, 1970). Though the monophyly of the genus *Monachus*, and therefore the monophyly of the tribe Monachini, has been questioned (Repenning and Ray, 1977; Wyss, 1988), this subfamily and tribal structure is widely accepted. Using multiple loci of varied evolutionary rates should allow phylogenetic resolution at these levels.

The second phylogenetic problem addressed is the species relationships within the Antarctic tribe Lobodontini. Morphological analyses using primarily cranial and dental characters (Hendey, 1972; de Muizon, 1982) placed the four species into two clades, *Lobodon+Hydrurga*, and *Leptonychotes+Ommatophoca*. Later study of cranial and some post-cranial characters (Bininda-Emonds and Russell, 1996) found the Lobodontini to be paraphyletic with respect to *Monachus*. This relationship is highly unlikely and probably artifactual. All molecular studies to date have supported *Hydrurga+Leptonychotes* as sister, but vary between placing either *Lobodon* or *Ommatophoca* as the most basal lineage (Davis et al., 2004; Fyler et al., 2005; Arnason et al., 2006; Fulton and Strobeck, 2006; Higdon et al., 2007). Supertree analysis combining molecular and morphological evidence (Bininda-Emonds et al., 1999) also recovered *Hydrurga+Leptonychotes*, but grouped the remaining two genera together as sister (*Ommatophoca+Lobodon*). Strong support for any of these hypotheses is rare (but see Arnason et al., 2006).

The third problem is one of both phylogenetic relationships and taxonomy. The northern seal tribe Phocini contains two subtribes (Chapskii, 1955a): Histriophocina (*Histriophoca, Pagophilus*) and Phocina (*Phoca, Pusa, Halichoerus*). While the monophyly of these subtribal

groups is well supported, neither use of molecular nor morphological evidence has been able to confidently resolve species relationships within Phocina. *Pusa* (*P. caspica, P. hispida, P. sibirica*) was historically considered a subgenus of *Phoca*, but was elevated by Scheffer (1958) to full generic status. Both *Pusa* (King, 1966; de Muizon, 1982) and *Phoca (Pusa)* (Chapskii, 1955a; McLaren, 1960a; Burns and Fay, 1970) have been applied since. For this study, *Phoca sensu stricto* includes two species, *P. vitulina* and *P. largha*. While morphological studies place *Pusa* and *Phoca* together, recent molecular work has indicated that the morphologically divergent *Halichoerus grypus* may belong within the *Phoca-Pusa* group, as opposed to its sister. Often, molecular studies place *H. grypus* either within *Pusa*, or as the sister to *Pusa*, to the exclusion of *Phoca sensu stricto* (Davis et al., 2004; Arnason et al., 2006; Fulton and Strobeck, 2006; Palo and Vainola, 2006; Higdon et al., 2007) (Figure 5-1). In particular, *H. grypus* is often recovered as the sister species of *Pusa caspica* (Arnason et al., 2006; Fulton and Strobeck, 2006; Palo and Vainola, 2006; Higdon et al., 2007). Although these Phocina species relationships are generally weakly or not supported, the consistent grouping of *H. grypus* within *Phoca sensu lato* (including *Pusa*) and the indication that the level of differentiation within Phocina is at the level of species differentiation compared to other phocids, not generic differentiation (Davis et al., 2004) has led to the suggestion that *Halichoerus, Pusa,* and *Phoca sensu stricto* all be included within the genus *Phoca* (Arnason et al., 1995; Arnason et al., 2006). In contrast, it has been suggested that *Halichoerus* is the only morphologically-warranted genus within Phocini, and that *Pusa,* and even *Histriophoca* and *Pagophilus,* are all sub-genera of *Phoca* (Burns and Fay, 1970).

In this study, I use multiple individuals per species and roughly 21, 000 base pairs of DNA sequence across various DNA marker types to obtain a comprehensive view of phocid systematics. By identifying model violations and selecting an appropriate partitioning strategy for mitochondrial DNA before combination with nuclear DNA, I am able to address taxonomy, provide a relatively novel hypothesis for Lobodontini evolution, and provide an important step towards reconciling morphological and molecular phylogenetic hypotheses for Phocina.

## Materials and methods

### *Sample selection, DNA isolation, amplification, and sequencing strategy*

47 taxa were included in this study. Seven non-pinniped carnivores were included as outgroups, comprised of two feliforms (*Felis catus*, cat; *Lynx canadensis*, Canadian lynx), two

canids (*Canis lupus*, grey wolf; *Alopex lagopus*, arctic fox), and three mustelids (*Mustela nivalis*, least weasel; *Meles meles*, European badger; *Enhydra lutris*, sea otter). When possible, two individuals from each phocid species were sequenced for each gene, and the most physically distant available samples were selected to best represent maximal diversity within the species. Individual sample and collection information is listed in Table 5-2. Due to sample availability, only one individual from *Pusa sibirica* and *Monachus monachus* and no representatives of *Mirounga leonina* were included in the nuclear data set and only one individual from *Monachus monachus* and *Mirounga angustirostris* were included in the mitochondrial dataset. Within each species, the same two individuals are not always represented, particularly between the mitochondrial and nuclear datasets due to use of some sequences from GenBank or to poor amplification and sequencing, though the same individuals are used across as many loci as possible. Table 5-3 lists the individual sequenced and the GenBank accession numbers for all loci. New sequences are available from the author upon request.

Total genomic DNA was isolated from either tissue or blood using the QIAgen DNeasy Tissue Extraction kit. Seventeen nuclear regions were selected (Table 5-4), representing 16 genes (2 fragments from *BRCA1*) and 15 unlinked regions (*RAG1* and *RAG2* are linked in the dog genome). Primer sequences are listed in Table 5-4. PCR amplification was generally performed in a 25μL total reaction volume, containing 0.4μM each primer, 160μM dNTPs, 1X PCR buffer (10mM Tris-HCl, pH 8.8, 0.1% Triton X-100, 50mM KCl, 0.16 mg/mL bovine serum albumin), 2.5mM $MgCl_2$, ~0.75U *Taq* DNA polymerase, and ~75ng genomic DNA. Cycling conditions for *BDNF*, *CHRNA1*, *FES*, *APP*, *GHR*, and *RHO* were: 94°C for 3 min, 30 cycles of 94°C for 30 sec., 54°C for 30 sec., and 72°C for 45 sec., with a final 5 min. hold at 72°C. The same conditions with annealing at 56°C were used for *PNOC*, *CREM*, and *PLCB4*. *ADORA3*, *APOB*, *RAG1*, *RAG2*, *BRCA1*.1, and *BRCA1*.2 conditions were: 95°C for 5 min, 35 cycles of 94°C for 60 sec., 56°C for 60 sec., and 72°C for 90 sec., with a final 5 min. hold at 72°C. *FLVCR1* was amplified using: 95°C for 5 min, 35 cycles of 94°C for 60 sec., 54°C for 60 sec., and 72°C for 120 sec.. $MgCl_2$ was reduced to 2mM for *BRCA1*.1, 1.2, and *FLVCR1*. *IRBP* was amplified in a total 100μL reaction with 1.75mM $MgCl_2$ with cycling conditions as per Stanhope et al. (1992). New mitochondrial gene sequences were obtained using the strategy and primers of Delisle and Strobeck (2002). Species-specific primers were designed when necessary. All PCR reactions were performed using an Eppendorf Mastercycler® ep. PCR amplification products were purified using the QIAgen PCR purification kit. All *IRBP* amplifications and other genes in some taxa could not be fully optimized and yielded more than one amplified region, as separated in a

1% agarose gel and visualized using ethidium bromide. In these cases, the PCR product of interest was excised from the agarose gel and isolated using the QIAquick gel extraction kit (Qiagen). Bi-directional direct sequencing was performed using the BigDye v3.1.1 (Applied Biosystems) chemistry per the manufacturer's instructions. The amplification primers were used for sequencing, with two loci (*IRBP* and *RAG1*) requiring additional internal sequencing primers (Table 5-4). Sequences were resolved using an Applied Biosystems 3730 capillary sequencer. Sequences were analyzed, basecalled, and aligned using the 3730 DNA Analyzer Data Collection Software (Foundation Data Collection) v.3.0, Sequence Navigator v.1.0.1, and ABIPrism SeqScape v.2.1 (all from Applied Biosystems). Heterozygous sites (two distinct peaks in the electropherograms observed from both directions of sequence) were coded as polymorphisms.

*Sequence alignment, model selection, base composition, and marker congruence*

Sequences were aligned by eye or using ABIPrism SeqScape and corrected by eye. Mitochondrial 12S rRNA alignment was performed using MAFFT v.6.240 (Katoh et al., 2002; Katoh et al., 2005) under the default settings for FFT-NSi (fast Fourier transform, iterative refinement) and adjusted manually in Se-Al v.2.0a11 (Rambaut, 2002) according to the carnivore 12S rRNA structural model of Ledje and Arnason (1996). When applicable, amino acid sequence was inferred from DNA sequence using Mesquite OSX v.1.12 (Maddison and Maddison, 2006). Phylogenetically informative insertion/deletion (indel) events were coded as present or absent, irrespective of length (Barriel, 1994) and their descriptions are included in Table 5-5. *BRCA1* fragments 1 and 2 represent different sections of the same exon and were thus concatenated for all analyses.

Base composition homogeneity across taxa was assessed in PAUP* for each dataset as a whole and partitioned by codon position. Base composition bias within each dataset as a whole was visualized in SeqVis v.1.3 (Ho et al., 2006), as well as partitioned by codon position. Congruence of each locus with the other loci in the same dataset was tested for both the nuclear and mitochondrial datasets using the incongruence length difference (ILD) test (Farris et al., 1995) implemented as the partition homogeneity test in PAUP* v.4.0b10 (Swofford, 2003).

The best-fit model of DNA evolution was selected for each partition separately, and for the combined nuclear and combined mitochondrial datasets, using the AIC criterion in MrModelTest v.2.2 (Nylander, 2004), which tests a restricted set of models from ModelTest (Posada and Crandall, 1998). Amino acid evolution models were selected using ProtTest v.1.4

(Abascal et al., 2005), which makes use of the PAL library (Drummond and Strimmer, 2001) and PhyML (Guindon and Gascuel, 2003).

*Tree search analyses*

The nuclear dataset was partitioned by gene and analyzed as both DNA sequence (nuclear dataset) and as a mixed set of DNA sequence for the non-coding loci and inferred amino acid sequence for the protein-coding genes (NuclearMixed dataset). The intronic locus *GHR* was determined to be incongruent with the other loci (see Results) and analyses were run both with and without *GHR*. All conclusions are based on analyses excluding *GHR*.

The mitochondrial (mt) DNA dataset was run using a number of different partitioning strategies, as described in the next section. The dataset was also analyzed as its inferred amino acid sequence plus 12S rRNA as DNA sequence.

Maximum parsimony (MP) analysis was performed on all datasets and all individual genes using PAUP*. Heterozygous sites were included as polymorphisms. Indel information was included in one analysis. MP analysis of the NuclearMixed dataset in PAUP* required coding the DNA segment as symbols 'AGCT', re-coding heterozygous bases as missing, and using a user-input stepmatrix, which was selected to be symmetrical and 'unweighted', where each change was equally probable. Tree searches were performed using 10 random addition starting trees and TBR branch swapping. Bootstrapping (BP) was performed under the same search parameters, using 1000 pseudoreplicates.

Maximum likelihood (ML) tree searching was performed using RAxML v.7.0.0 (Stamatakis, 2006a, b). Each gene partition was allowed its own parameter set within the GTR+Γ model for both tree searching and rapid bootstrapping (100 replicates, GTR+CAT approximation model). The JTT transition matrix used for all protein-coding genes in the NuclearMixed dataset, as selected by ProtTest. MtMam was used for all mt protein-coding genes, except COI, where MtRev was applied. Mt amino acid (aa) models were determined from the most frequently sampled matrix in MrBayes when aa model was allowed to vary (below).

Bayesian inference (BI) was performed in MrBayes v.3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Altekar et al., 2004). For all partitioned analyses, base frequencies, transition:transversion ratio, gamma shape, proportion of invariance, and amino acid model were unlinked across partitions, where applicable. Rates were allowed to vary across partitions (prset=variable), using a flat Dirichlet prior. For the NuclearMixed dataset, the amino

acid models were sampled as additional parameters (aamodelpr=mixed). Two simultaneous runs of 4 chains each with the default heating parameter of temperature=0.2 were performed. For the NuclearMixed dataset, chain mixing was poor, thus, heating was removed and 8 independent chains were run. Convergence was assessed and an appropriate burn-in period selected using both the statistics provided in MrBayes (parameter PSRF≈1.0, using *sump* and average standard deviation of split frequencies between runs <0.01) and visualizing the trace plots and comparing parameter values and distributions between runs in Tracer v.1.4 (Rambaut and Drummond, 2007). Each analysis varied in the number of generations required before convergence both within and between runs could be achieved. These ranged from 1 to 10 million generations. All trees were visualized in TreeView X (Page, 1996).

*Mitochondrial partitioning strategies*

Multiple partitioning strategies of the mtDNA dataset were implemented in MrBayes. These included use of a single partition, partitioned by gene using the best-fit models as selected by MrModelTest (13 partitions), by codon position (4 partitions: Pos1, Pos2, Pos3, 12S rRNA), by both codon position and gene (37 partitions) and by rate (6 partitions). Rate was inferred indirectly from character consistency indices, following Kjer and Honeycutt (2007). To determine these 'rate' partitions, 1000 MPBP replicates were first performed in PAUP*, saving one tree per replicate. Characters were then weighted by maximum fit CI and character status (cstatus) was copied into Microsoft Excel and sorted by weight. Six categories were assigned based on natural breaks in the weight classes and attempting to keep the categories similarly sized. The Slowest category contained 7603 characters (base pairs, bp) of weight=1.0; Slow: 522 bp, weight=0.6-0.75; Med: 1114bp, weight=0.4285-0.5; Med_Fast: 937bp, weight=0.3-0.4; Fast: 1039 bp, weight=0.2-0.2857; Fastest: 571 bp, weight=0.09-0.181.

Bayes factors (BF) were used to investigate which partitioning strategy for mtDNA was preferred. Bayes factors are the ratio of the marginal likelihoods of two competing models ($H_1$, $H_2$). Models do not have to be nested and a more complex model is not necessarily preferred. There is no defined significance level, but the test statistic $2\ln(BF_{21})$ with a value >10 is generally used to indicate "very strong" evidence against $H_1$ (Kass and Raftery, 1995) and this cut-off appears to perform relatively well for empirical data (Brown and Lemmon, 2007). Since marginal likelihoods are difficult to compute, the harmonic mean (HM) estimator is used (Newton and Raftery, 1994). The natural logarithm of the harmonic mean was calculated from

the stationary phase MCMC samples using *sump* in MrBayes. The test statistic is then calculated as $2(\ln HM_2 - \ln HM_1)$ (Nylander et al., 2004; Brandley et al., 2005; Brown and Lemmon, 2007).

*Partition exclusion to avoid base composition bias across taxa and RY coding*

Significant base composition bias across taxa was observed (see Results) for the 3rd codon position in the MtDNA dataset. Therefore, the 3rd codon positions were removed and the 1st and 2nd positions were analyzed using the minimum evolution (ME) criterion with LogDet distance (Lockhart et al., 1994), which is more resistant than other models to any potentially remaining violations of the assumption of base composition stationarity. MLBP in RAxML was also performed for this dataset as a single partition. In addition to the above described mt aa analyses, the dataset was also recoded to purines and pyrimidines (RY coding) and analyzed using MP as above.

*Combined nuclear and mitochondrial analyses*

Based on the mt partitioning strategy selection results (see Results), the nuclear dataset was combined with mtDNA partitioned by rate. The nuclear dataset was also combined with the mtDNA with the 3rd codon positions excluded. For these two combined datasets, BI in MrBayes and MLBP in RAxML were performed as described above. BI was also performed for a combined dataset of nuclear DNA and mtProt.

*Topological effects of multiple individuals and inter- and intraspecific variation*

The effect of including multiple individuals from each species on the recovered topology was examined by randomly excluding different sets of individuals, such that each species was only represented by a single taxon. MP and MPBP searches were performed in PAUP*.

The level of variation within and between species was examined by calculating the distance between each pair of taxa in PAUP*. LogDet distance was used for mtDNA and GTR distance was used for the nuclear dataset. Taxonomic comparisons were made between species, genera, tribes, subfamilies, and families. Lower taxonomic levels were averaged when a higher comparison was made to that groupings with more taxa were not given a higher weight. For example, when Phocinae tribes were compared, an average distance for each Phocina genus to

*Cystophora* was calculated separately, then all genus averages were combined to calculate the average Phocini vs. Cystophorini distance.

## Results

### *Sequencing results*

Amplification could not be achieved, despite numerous changes to the PCR conditions for *RAG2* in the two canids, *BRCA1*.1 in *O. rossii*, *BRCA1*.2 in *L. canadensis*, *IRBP* in one individual of each *E. jubatus, L. weddellii,* and *M. schauinslandi*, and *FES* in one individual of *L. weddellii*, and *H. leptonyx*. *FES* sequence was not included for *F. catus*, because it could not be confidently aligned. A 221 base pair (bp) insertion in *FLVCR1* was removed for both *Lynx* and *Felis*. This insertion is most likely a SINE belonging to the family of feliform SINEs described by Pecon Slattery et al. (2000), including a $(CT)_n$ (where n=8 or 11 in these individuals) region followed by a poly-A tail. It is inserted in reverse orientation to the coding direction of the gene. A poly-A region, ranging from 6-14 bp, was observed in all individuals in *CREM* and often could not be successfully sequenced through due to slippage. Therefore, many *CREM* sequences are single-stranded on either side of the poly-A. *CHRNA1* in *P. sibirica*2 and *FLVCR1* in *P. groenlandicus*1 and both *O. rossii* individuals were observed to be heterozygous for an allele containing a deletion. The longer allele was included as the sequence, but the sequences were single-stranded on either side of the deletion site.

One region of 12S rRNA corresponding to a section of HVR9 (Ledje and Arnason, 1996) could not be confidently aligned and was excluded. The complete aligned length of the nuclear dataset was 9568 bp and the mt data set was 11794 bp (Table 5-6). When the protein-coding genes were converted from DNA to amino acid (aa) sequence, the mt dataset was 3619 aa + 961 bp (12S rRNA) and the NuclearMixed dataset was comprised of 2085 aa + 3295 bp.

### *Base composition and congruence analyses*

The nuclear dataset did not show any base composition bias among taxa (p=0.9999). The mtDNA dataset did show significant base composition bias across taxa (p=0.0000), but no bias was observed after the 3$^{rd}$ codon position bases were removed (p=0.9999). Combining the nuclear and mt datasets together showed significant bias across taxa (p<0.0001) whether the mt

3rd positions were included or not. Visualization of the base composition of each the mt and nuclear datasets as a whole indicated that there was very little bias in the nuclear dataset, but large bias in the mt dataset (Figure 5-2 a, b). Both nuclear and mitochondrial protein-coding genes show different base composition across codon positions (Figure 5-2 c, d). Low spread of the data points (representing taxa) for all nuclear codon positions and mitochondrial 1st and 2nd positions indicate base composition stationarity across taxa. Increased spread of the mt 3rd position data points is consistent with the finding in PAUP* that base composition is heterogenous across taxa.

ILD testing for congruence between each gene and all other genes found 5 nuclear genes to be significant at $p \leq 0.05$, of which *GHR* was significant at $p \leq 0.01$ (Table 5-6). One mt gene (ND5) had a p-value of 0.03 and another (ND3) was 0.01. Due to the tendency of the ILD test toward Type I error (rejecting a true null), a significance level of $p=0.05$ is too high (Sullivan, 1996; Cunningham, 1997), thus, a significance level of 0.01 is used to reduce Type I error. There are several problems with use of the ILD test, particularly the effects of phylogenetic noise, relative partition contribution, and heterogeneity in evolutionary rate (Dolphin et al., 2000; Barker and Lutzoni, 2002; Darlu and Lecointre, 2002; Dowton and Austin, 2002; Quicke et al., 2007). It has been suggested that the ILD test never be used for measuring partition combinability (Yoder et al., 2001; Barker and Lutzoni, 2002), so further examination of the 'rejected' partitions was performed.

Maximum parsimony bootstrap (MPBP) gene trees were examined for strongly supported (>80%) incongruent clades to determine if topological differences exist. The ND3 gene tree strongly supports a non-monophyletic *Histriophoca*. Removal of the 3rd codon position for this gene results in the partition no longer being significantly different from the other genes and MPBP of ND3 1st and 2nd positions finds no support against *Histriophoca* as monophyletic. Inclusion/exclusion of ND3 from the mt dataset had no effect on the topology recovered or on its support. ND3 3rd positions may be incongruent due to noise, and as it does not appear detrimental to phylogenetic reconstruction in the dataset as a whole, it is not excluded from the mtDNA dataset. The *GHR* MPBP tree did not strongly support any different clades from the remaining gene trees, but did show moderate support (69% MPBP) grouping both individuals of *P. caspica*, both *H.grypus*, one *P. hispida*, and one *P. vitulina* together, which could potentially be driving the incongruence. Tree searches were performed both with and without *GHR*. For ML analysis of DNA, no topological changes were observed, but support for clades within *Pusa* increased from 26% MLBP to 45% (*P. caspica* + *P. sibirica*) and 39% to 63% (*Pusa* monophyly), while support for *Phoca largha* species monophyly was reduced from 86% to 49%. ML analysis of the

NuclearMixed dataset with and without *GHR* showed a similar but smaller effect, with the aforementioned MLBP clade support changing from 34% to 48% (*P. caspica* + *P. sibirica*), 54% to 63% (*Pusa*) and 82% down to 59% (*Phoca largha* monophyly). MP analysis of the nuclear dataset accordingly showed MPBP changes from 17% to 31% and 38% to 56% within *Pusa*, but recovers *Phoca largha* as paraphyletic with respect to *Phoca vitulina*, grouping one *P. largha* individual with *P. vitulina* (*P. vitulina* monophyly is strongly supported) with very weak support (MPBP=46%). Because of these topological effects on the combined nuclear topology and the significant ILD result, conclusions and figures are based on those searches excluding *GHR*.

Interestingly, comparing mt and nuclear DNA partitions does not result in rejecting the null hypothesis of congruence (p=0.84). The partitions differ topologically in areas that are generally very weakly supported and/or unresolved, so the congruence is potentially true. This is compounded by the large contribution of the MtDNA datset to the tree score. It may be that the nuclear dataset does not impact the tree score to a large enough extent to affect the sum of the tree scores from the separate partitions, thus, partition size could be greatly affecting significance (Dowton and Austin, 2002).

*Nuclear phylogenetic results*

All families, subfamilies, and tribes were strongly supported as monophyletic by all analyses (Figure 5-3, Table 5-7). Lack of resolution or low support was generally restricted to species relationships within Lobodontini and Phocini. All nuclear analyses of either the DNA or NuclearMixed dataset recovered the same topology, with varying levels of support (Figure 5-3, Table 5-7). Within Lobodontini, two clades are recovered. *Leptonychotes* + *Hydrurga* was strongly supported (BPP=1.0, ML and MPBP ≥99%), while *Ommatophoca* + *Lobodon* was weakly to moderately supported (BPP=0.95, MLBP=54,61%, MPBP=66-67%). Within Phocini, monophyly of both subtribes, Histriophocina (BPP=1.0, MLBP≥98%, MPBP≥96%) and Phocina (BPP=1.0, MLBP≥91%, MPBP≥86%), was supported. Within Phocina, monophyly of *Phoca sensu stricto* was strongly supported, and though both *Phoca* species were recovered as monophyletic, support for *P. largha* monophyly was low when *GHR* was excluded (see above). MPBP analysis of the Nuclear dataset recovered the only different topology from all other methods, recovering *P. largha* as paraphyletic with respect to *P. vitulina*. *Pusa* + *Halichoerus* was moderately supported (BPP=1.0, MLBP=70-78%, MPBP=67,74%). *Pusa* is consistently recovered as monophyletic with lower support (BPP=1.0, MLBP=60-68%, MPBP=58,64%).

*Pusa sibirica* and *P. caspica* are consistently recovered as sister, with moderate Bayesian posterior probability (BPP=0.95), but no bootstrap support greater than 50% (MLBP=40-48%, MPBP=30, 36%).

BI of the NuclearMixed dataset did not fully reach convergence either using the default temperature or with heating removed. For the analysis with no heat, two 'sets' of chains were observed and the four chains that had converged on the same distribution were analyzed. The average standard deviation between chains was <0.001, but the PSRF of all parameters was not 1.0. Both analyses (with and without heating) recovered very similar clade support values (BPP=0.99-1.0 for all nodes except *P. largha* monophyly) and the same maximum *a posteriori* (MAP) tree. Both recovered the topology shown in Figure 5-3 with BPP~0.8 and a second topology where *P. largha* was paraphyletic with respect to *P. vitulina* with BPP~0.2. BI of the nuclear dataset recovered the same two trees (BPP=0.533 and 0.357) and additionally recovered a third tree within the 90% confidence interval (BPP= 0.036) placing *Ommatophoca* basal within the Lobodontini. It is of note that both the MAP tree and the second tree in the 99% CI for the NuclearMixed dataset include the sister relationships between *Ommatophoca* + *Lobodon* and *Pusa sibirica* + *Pusa caspica* and both clades had BPP=1.0. Difficulty in achieving convergence for some parameter estimates in the NuclearMixed dataset is primarily attributed to the small partition sizes for some of the loci (see Table 5-6).

*Mitochondrial phylogenetic results*

Analysis of the mtProt and mtDNA datasets also recovered strong support for all higher-level groupings, but different analyses recovered different relationships within Phocina and Lobodontini (Figure 5-4, Table 5-8). All likelihood and Bayesian analyses of the entire mt dataset recovered *Pusa* as paraphyletic with respect to *Halichoerus*, where *Halichoerus* is sister to *Pusa caspica*. Placement of *P. sibirica* and *P. hispida* varied (Figure 5-4 a-d), although partitioning by rate showed low support for *P. sibirica* as sister to *Halichoerus-P.caspica* (Table 5-8). MPBP recovered monophyletic *Pusa* as sister to *Halichoerus* (Figure 5-4 b). MPBP also recovered the unusual branching of *Histriophoca* before *Pagophilus* with weak support (MPBP=58%), rendering Histriophocina paraphyletic. This result is not observed in any other analysis. In all areas of conflict, support is low (Table 5-8). MtDNA analyses generally showed good support for the *Halichoerus-Pusa* clade, to the exclusion of *Phoca* and for the monophyly of Phocina, Histriophocina, and Phocini. Within Lobodontini, all analyses of the mtProt dataset

recovered *Ommatophoca* as basal, then *Lobodon* with low MLBP support, but strong Bayesian support (Table 5-8), then a strongly supported *Leptonychotes* + *Hydrurga* clade (Figure 5-4 g). This topology was also recovered by MPBP (one partition) and BI partitioned by codon position, but placement of *Ommatophoca* and *Lobodon* is reversed when mtDNA is either partitioned by rate or analyzed using ML or BI as a single partition (Figure 5-4 e).

Exclusion of all 3rd codon position bases (i.e. including only 12S rRNA, 1st and 2nd positions) and analysis under the minimum evolution criterion with LogDet distance recovered the same topology as the Nuclear dataset for both Phocina and Lobodontini (Figure 5-4 b, f). Monophyletic *Pusa* is well-supported (MEBP=75%), with *P. caspica* and *P. sibirica* recovered as sister (MEBP=52%). In Lobodontini, *Lobodon* and *Ommatophoca* were weakly supported as sister (MEBP=57%). MLBP analysis excluding the mt 3rd position in RAxML recovered *Pusa* + *Phoca* with moderately strong support (MLBP=82%), with *Halichoerus* outside the clade. Within *Pusa* + *Phoca*, there was no support (>50%) for the branching order, except monophyly of *Phoca* (MLBP=100%). As in the LogDet analysis of this dataset, *Lobodon* and *Ommatophoca* were recovered as sister (Figure 5-4 f), but it was unsupported (MLBP=32%).

*MtDNA partitioning strategies*

Partitioning the mtDNA dataset by gene (BF=1468.86), by codon position (BF=10828.74) or by rate (BF=19039.56) were selected over not partitioning. Partitioning by rate was selected over either partitioning by gene (BF=17570.70) or by codon position (BF=8210.82). Despite several changes to the Bayesian analysis, convergence could not be achieved for the dataset partitioned by both gene and codon position. Perhaps because convergence could not be achieved, a lower –lnL was obtained for this dataset than either partitioning by rate or by position alone. Both rate (BF=8532.68) and codon position (BF=321.86) partitioning strategies were preferred over both gene and position, but partitioning only by gene was not (BF=-9038.02). Because partitioning by rate was strongly selected over all other partitioning strategies tested, it was included in the combined mt and nuclear analyses, as was the mt dataset excluding 3rd positions.

MP analysis of the RY-coded dataset did not yield much resolution for the tip clades and is thus not further discussed.

*Combined mt and nuclear results*

Combining the nuclear and mtProt datasets recovered similar results as analyzing mtDNA alone, but with some reduced support within Lobodontini and no support (BPP=0.71, NuclearMixed + mtProt) for the recovered *P. caspica* + *H. grypus* clade, nor *P. sibirica* + *P. hispida* (BPP=0.79).

ML analysis combining the nuclear genes and mt without 3rd positions recovers a novel topology (Figure 5-5 a). Similar to the ML analysis excluding the 3rd positions, *Halichoerus grypus* is most basal within the Phocini, followed by *P. vitulina* + *P. largha*, then *Pusa*. Both *Pusa* and *Phoca*+*Pusa* were recovered with moderate support (MLBP=71%, 74% respectively). BI also recovered monophyletic *Pusa*, but as sister to *Halichoerus* (unsupported, Figure 5-5 a). Two clades were recovered within Lobodontini, as in the nuclear analyses, with *Ommatophoca* + *Lobodon* weakly supported (Figure 5-5 c: BPP=0.95, MLBP=57%). When the nuclear dataset was combined with the entire mtDNA partitioned by rate, *Pusa* was found to be paraphyletic with respect to *Halichoerus* as in mtDNA analyses, though it is very poorly supported (Figure 5-4 b) while the topology for the Lobodontini (Figure 5-4 c) is the same as the nuclear analyses.

*Use of multiple individuals per species*

Analyses using different sets of individuals, representing only one individual per species generally did not have a large effect. MPBP values did not vary largely between datasets, with some clade support increasing and some decreasing. However, topological changes within the Phocini were observed in both the nuclear and mt datasets. MPBP analysis including 2 individuals per species in the nuclear dataset did not support any clades >50% within *Pusa* + *Halichoerus*, except for the monophyly of each species. Use of only *H. grypus1*, resulted in *P. caspica* grouping with *H. grypus*, either weakly (MPBP=51%) or moderately (MPBP=77%), depending on which *Pusa* individuals were included. Using *H. grypus2* instead resulted in either a polytomy of all *Pusa* species + *H. grypus* or recovered a moderately supported *P. caspica* + *P. hispida* clade (MPBP=71-76%). In the mtDNA dataset, MPBP analysis using 2 individuals per species recovered a monophyletic *Pusa* (MPBP=63%), with sister *H. grypus*. When only one individual per species is used, regardless of the combination of individuals, the two *Phoca* species are drawn into the *Pusa* + *Halichoerus* clade with moderate support (MPBP=61-76%). No other topological changes were observed, though certain combinations of individuals resulted in

reduced MPBP support for *Ommatophoca* branching first within the Lobodontini using mtDNA (MPBP~75% to MPBP=58-76%).

*Inter- and intraspecific variation*

Pair-wise distances between taxa were ~15-20 times lower for the nuclear dataset than the mt dataset, using LogDet distance for both. GTR and LogDet distance measures were very similar for the nuclear dataset, but LogDet distances were slightly higher for close relationships and lower for very distantly related taxa in the mt dataset. A number of taxonomic comparisons were made and the average distances between the taxa being compared are visualized in Figures 5-6 (mt) and 5-7 (nuclear). Comparisons were made within species (between individuals), within genus (between species, with individuals averaged per species), pair-wise between genera of the same subtribe for Phocina and Histrophocina (between species, with individuals averaged per species), between subtribes (pair-wise genus comparisons, with species averaged per genus), and pair-wise between tribes (genera averaged within subtribe). Distances were also calculated at higher taxonomic levels but not shown in the figures. Mt LogDet distances between the three pinniped families were roughly equal (~0.2) and were at the low end of the range of distances of comparisons between all carnivore families (~0.21-0.27). Nuclear GTR distance from Phocidae to either Otariidae or Odobenidae was ~0.02-0.025, while distance between Otariidae and Odobenidae was ~0.012, which is within the range of distances observed between species belonging to the different Phocidae subfamilies (~0.010-0.014). Distance from the pinniped families to Mustelidae was ~0.075 and to canids and felids was ~0.085. It is important to note that all of these distances, particularly those within species, are not corrected in any way for proximity of the individuals sampled within the species range, population size or structure or any other factors that may affect intraspecific genetic distance.

**Discussion**

*Partitioning, model use, and conflicting datasets: How to choose?*

Mitochondrial (mt) sequence data has historically dominated molecular phylogenetic analysis; only recently have large-scale nuclear studies become increasingly common. Many factors including low recombination rate, high mutation rate, and the comparative ease of marker

development and sequencing have made mtDNA a widely-used phylogenetic marker. However, this increased mutation rate more quickly leads to saturation and difficulty in resolving deep relationships. Base composition bias is common and composition can differ across lineages, potentially leading to artifactual taxon groupings. Though the mitochondrion represents a single hereditary unit, different genes or sections may be under different selection pressures or have different mutational rates. All of these problems increase difficulty in proper modeling, and therefore, reliable tree estimation. Therefore, I examined multiple strategies of analyzing the mt dataset.

Bayes factors (BF) were used to decide between alternate partitioning schemes of the complete dataset. Any partitioning scheme was strongly preferred over analysis of mtDNA as a single partition. Mt partitions were determined using two different strategies. The first is the traditional partitioning based on gene and/or codon position. The second partitioning strategy accounts for heterogeneity in nucleotide site evolution differently by use of rate categories (Kjer and Honeycutt, 2007). This method assigns each nucleotide to a user-defined number of categories based on their consistency index (CI). This effectively partitions characters roughly by their rate, as the characters with the lowest CI are inferred to have the fastest rate and vice versa. Bayes factors selected partitioning by rate over all other strategies tested. Partitioning by codon position was selected over both partitioning by gene and by both gene and codon position. Convergence could not be achieved for the dataset partitioned by both codon position and gene, so the preference for partitioning only by position (BF=321.86) cannot be definitive. However, partitioning by rate was very strongly preferred (BF=8532.68) and it is unlikely that even with full convergence that partitioning by both gene and rate would be preferred. Given the difficulty in obtaining convergence, the highly partitioned by-gene-by-position strategy is likely overparameterized and not ideal for this particular dataset.

Overpartitioning, and, thus, overparameterizing, can be problematic (Rannala, 2002; Lemmon and Moriarty, 2004), though underpartitioning can be worse (Brown and Lemmon, 2007). Although very few topological changes were observed, strong evidence against both the least and most highly partitioned datasets suggest that both of these problems could be occurring with the strategies employed for this empirical data. Very strong BF evidence for any partitioning strategy follows recent empirical studies in recovering values 10 to $10^3$ times higher than the "very strong evidence" level of 10 (Mueller et al., 2004; Nylander et al., 2004; Brandley et al., 2005; Castoe and Parkinson, 2006; Clayton et al., 2007). But in contrast to many of these studies, the most highly partitioned model (by gene and by position: 37 partitions) was not

selected, instead finding the greatest support for the more "statistical" method (CI based) of partitioning into rate classes, as opposed to more "biological" (by codon position or gene) partitioning strategies. Different partitioning strategies did recover different topologies for the two poorly resolved regions of the tree (Figure 5-4). These results emphasize the importance of properly selecting a partitioning strategy that best fits the data in hand, not simply selecting either the most highly partitioned or unpartitioned method without examination.

An alternative strategy of partitioning data is to simply remove partitions that either violate model assumptions or are inconsistent with other partitions. The 3rd codon position in mitochondrial genes has been long recognized as problematic in phylogenetic studies due to its fast mutation (and therefore saturation) rate and often for differing base composition across lineages. This study is no exception, as the 3rd position bases were found to be significantly heterogenous in their base composition across taxa, both statistically in PAUP* and visually in SeqVis (Figure 5-2). When all mt 3rd positions are removed, no base composition bias across taxa is observed. Analyses of this dataset become consistent with nuclear DNA results (Figure 5-3, 5-4 b, f), though ML analysis recovers *Phoca* and *Pusa* together (to the exclusion of *Halichoerus*) consistent with morphology.

Given that including all mtDNA and partitioning by rate or excluding mt 3rd positions yield different topologies, can one be determined to be more reliable? Because we cannot know the true tree, these recovered topologies must be compared to topologies recovered from other independent markers. Combining multiple nuclear loci with varying evolutionary rates yielded a very consistently recovered topology and high consistency indices for all included loci (Table 5-6). If consistency with these nuclear markers and morphology is the criterion, then excluding mt 3rd positions is the preferred solution for this dataset. Any likelihood-based analysis including all of the mtDNA recovers *Pusa* as paraphyletic with respect to *Halichoerus*, a result that is inconsistent with the nuclear results and highly inconsistent with morphology. In contrast, LogDet analysis of the dataset excluding mt 3rd positions recovers the same topology as the nuclear analyses, while ML analysis recovers a topology for the Phocina that is more consistent with morphology (*Phoca+Pusa*). Thus, though both results are discussed, excluding the partition that violates model assumptions of base composition stationarity appears to yield a more reliable topology.

*Higher-level phocid relationships*

The superfamily Otarioidea (Otariidae + Odobenidae) has a long history of morphological support extending back to Brookes (1828) and Gill (1866), which led to Allen's (1880) division of phocids and otarioids into "wrigglers" and "walkers" (see Scheffer, 1958; Wyss, 1988). Otarioidea is strongly supported by all analyses, including the presence of four 3-12 bp indels (Table 5-5). This provides evidence that their grouping in molecular studies is not a long branch artifact, as has been proposed to explain the difference between some morphological and molecular studies (Berta et al., 2006). Otariidae and Odobenidae show a lower genetic distance to one another than either does to Phocidae based on nuclear DNA. However, all three families are relatively equidistant based on mtDNA, supporting Odobenidae as its own family.

The division of Phocidae into two subfamilies (King, 1966) and six tribes (*sensu* Burns and Fay, 1970) was also strongly supported by both nuclear and mt analyses. These classifications are also reflected in the genetic distance between taxa from different tribes, genera, or species. There is a close relationship between mt genetic distance and taxonomic 'distance' (Figure 5-6), though the relationship using nuclear loci is considerably less clear (Figure 5-7). In both marker types, genetic variation at any given taxonomic level is always lower than that of the higher taxonomic level within a lineage.

Within Phocinae, Erignathini (*Erignathus barbatus*) represents a very early divergence (Figure 5-3, 5-4). Particularly for nuclear DNA, the distance from Erignathini to either other phocine tribe is much higher than Cystophorini and Phocini are to one another (Figures 5-6, 5-7 i). A deep divergence is in accordance with morphological and karyotypic studies that have noted some of the 'monachine-like' tendencies of *Erignathus*, supporting it as either an intermediate or link between the subfamilies (King, 1966; Fay et al., 1967; Burns and Fay, 1970).

Within Monachinae, genetic distance between tribes is more similar, particularly at the nuclear loci (Figures 5-6, 5-7 n). However, the distance between *Monachus* species (Figures 5-6, 5-7 l) represents the most obvious exception to the correlation between genetic and 'taxonomic' distance. Genetic distance between the Hawaiian monk seal (*M. monachus*) and the Mediterranean monk seal (*M. schauinslandi*) is on the order of 'tribal-level' differentiation, not 'species-level' (Figures 5-6, 5-7). The two species share their common ancestor at a very deep node in the tree (Figure 5-3 node 1, Figure 5-4 node 7), so a high level of genetic divergence is unsurprising. However, the monk seals are very conserved morphologically, resembling extinct relatives in their primitive appearance (Hendey, 1972; Repenning and Ray, 1977; Wyss, 1988) in

contrast to their high genetic divergence. The two species are very strongly supported as sister, confirming that all higher-level designations (Table 5-1) are appropriate based on both the recovered phylogeny and the relative divergence levels.

*Phocina species relationships*

One area of topological conflict between different marker or analysis types occurs within Phocina, a subtribe of the Phocini. Phocini and both of its subtribes, Phocina and Histrophocina, are strongly supported as monophyletic and only the species relationships within Phocina are problematic (Tables 5-7, 5-8). All analyses of nuclear loci consistently recovered the same Phocina topology (Figure 5-3), but with varying levels of support (Table 5-7). A monophyletic *Pusa* (*P. caspica, P. sibirica, P. hispida*) was recovered as sister to the monotypic genus *Halichoerus*, and *Pusa-Halichoerus* as sister to *Phoca*. In contrast, most mt analyses, including amino acid analysis or when mtDNA is partitioned by rate, placed *Halichoerus* within *Pusa* (Figure 5-4), rendering the latter paraphyletic, even when combined with nuclear genes (Figure 5-4b). Like in the nuclear analyses, MP analysis of mtDNA (Figure 5-4 c), and all analyses when mt 3rd positions are excluded (Figure 5-4 b, 5-5) recover *Pusa* as monophyletic. This is strongly preferred by morphology over the inclusion of *Halichoerus*, suggesting that the placement of *Halichoerus* within *Pusa* by most mt analyses here and in other recent investigations primarily utilizing mitochondrial DNA (Arnason et al., 2006; Fulton and Strobeck, 2006; Palo and Vainola, 2006; Higdon et al., 2007) is artifactual. *Halichoerus* is on a comparatively long branch when recovered inside *Pusa* by mt aa analysis or combined nuclear + mtDNA (by rate), which may also indicate that the paraphyly of *Pusa* is artifactual. Because this paraphyletic result is not observed when the 3rd codon positions are removed, it may be due, at least in part, to the significant base composition bias across taxa at the mt 3rd codon position.

In nuclear, mtProt, and most mtDNA analyses, *Phoca sensu stricto* is recovered outside the *Halichoerus-Pusa* clade (Figures 5-3, 5-4). However, MLBP analyses (but not BI) of the mtDNA with 3rd positions excluded, either alone (82%) or combined with the nuclear data (74%), supported *Pusa* and *Phoca* as sister (Figure 5-5a). While *Pusa+Halichoerus* was recovered and supported by more analyses, *Pusa+Phoca* is more consistent with traditional morphological analyses. Because both can be moderately supported, but neither consistently strongly supported, it cannot presently be determined which represents the 'true' topology. *Halichoerus* is not very genetically distant from either the *Pusa* or *Phoca* species. Genetic distance between *Halichoerus*

and any of the *Pusa* species is roughly equivalent to any species comparison within *Pusa* (Figures 5-6, 5-7 c, d). The average distance between *Halichoerus* and *Pusa* is somewhat lower than either is to *Phoca* using nuclear DNA (Figure 5-7, d vs. e, f). Other genera within Phocinae are much more genetically distant from one another (Figures 5-6, 5-7 g, h), as are those within the Monachinae (Figures 5-6, 5-7 m), with the exception of the comparatively low genetic distance between lobodontine sister genera *Leptonychotes* and *Hydrurga* (Figures 5-6, 5-7). Thus, whether *Halichoerus* is resolved as the sister to *Pusa* or to *Pusa-Phoca*, it shows high morphological divergence compared to the very low genetic distance. From a strictly genetic perspective, there is no threshold that can be used to delimit genus-level differentiation within the closely related Phocini.

Within *Pusa*, nuclear analyses recover the two land-locked *Pusa* species (*P. sibirica* and *P. caspica*) as sister. This clade was recovered by all analysis methods and supported by BI, but not supported by either MLBP or MPBP (Table 5-7). Given the low BP support, it is probable that the Bayesian clade posterior probabilities (BPP) are inflated for this very short, potentially polytomous branch (Lewis et al., 2005). However, the posterior probability of the topology, as opposed to clade BPP, is sometimes considered as a better representation of the optimal tree (Wheeler and Pickett, 2008). For both the nuclear and NuclearMixed analyses, the maximum *a posteriori* (MAP) tree recovered *P. caspica* and *P. sibirica*. So, though clade bootstrap support values are not high, consistent recovery by all analysis methods indicates a certain level of confidence that the two land-locked *Pusa* species could be true sister taxa. However, combined analyses excluding mt 3$^{rd}$ codon positions group *P. sibirica* with *P. hispida*, but the relationship is not supported (Figure 5-5 a). This grouping is also sometimes observed for mitochondrial analyses (Figure 5-4 c, d), but again, is not supported. Unfortunately, given the high morphological similarity between species, there is no morphology-based consensus on the relationships within *Pusa*, though most studies recover one of the two relationships as in this study. Rarely are *P. caspica* and *P. hispida* proposed as sister (for review, see Palo and Vainola, 2006). Though species relationships within *Pusa* cannot be confidently resolved, it is an important step that the genus is supported as monophyletic. Molecular studies until now have often recovered a paraphyletic *Pusa* (Figure 5-1), but generally with unsupported or polytomous relationships. Consistent recovery of a monophyletic *Pusa* using nuclear or combined data in this study shows promise that increasing nuclear markers could yield a better supported topology that is more compatible with morphology for all Phocina species.

*Lobodontini species relationships*

Within the Antarctic tribe Lobodontini, both mt and nuclear analyses strongly supported *Leptonychotes weddellii* (Weddell seal) and *Hydrurga leptonyx* (leopard seal) as sister. They are also considerably less genetically divergent from one another than any other genus-level comparison within Monachinae (Figures 5-6, 5-7 l). Although this sister relationship was not found by morphological studies (Hendey, 1972; de Muizon, 1982), it has been consistently strongly supported by molecular studies (Bininda-Emonds et al., 1999; Davis et al., 2004; Fyler et al., 2005; Arnason et al., 2006; Fulton and Strobeck, 2006; Higdon et al., 2007). These two species do share some superficial morphological similarity, such as their spotted coats, and *L. weddellii* was originally described by James Weddell as the "sea leopard" (Scheffer, 1958). It has also been referred to as the "false sea leopard" (see Hince, 2000) to differentiate from the true 'sea leopard', the leopard seal. *Hydrurga leptonyx* are highly predatory, often preying on other seals (primarily crabeater seals), but also rely heavily on krill (Rogers, 2002). *Lobodon carcinophagus*, the morphologically-defined sister taxon to *H. leptonyx*, relies almost solely on krill (Bengtson, 2002), despite its common name, the crabeater seal. Both *H. leptonyx* and *L. carcinophagus* have specialized dentition for this food source. Given the strong molecular support for the sister grouping of *H. leptonyx* with *Leptonychotes weddellii*, as opposed to the morphological grouping with *Lobodon carcinophagus*, re-evaluation of morphology with a decreased focus on dental characters would be quite useful. It is possible that adaptations for eating krill are convergent between the latter two species and that adaptations to different li*FES*tyles have obscured other morphological similarities between *H. leptonyx* and *L. weddellii.*

The placement of the remaining two species of lobodontine seals is less resolved. Mt analyses, including partitioning by rate, recover either *Lobodon carcinophagus* or *Ommatophoca rossii* as the basal lineage (Figure 5-4, 5-4d), but always with low support (Table 5-8). Excluding mt 3$^{rd}$ codon positions (Figure 5-4 f), nuclear analysis (Figure 5-3), or combined nuclear and mt (Figure 5-5 c), recover *O. rossii* and *L. carcinophagus* as sister, but always with low or no support (Table 5-7, 5-8, Figure 5-5). A close relationship between *L. carcinophagus* and *O. rossii* has previously only been observed in a supertree combining morphological and early molecular work (Bininda-Emonds et al., 1999). Consistent recovery of this clade suggests that this may be a valid hypothesis for their evolution that deserves further consideration and scrutiny.

*Use of multiple individuals*

Though the inclusion of multiple individuals per species generally had no effect on the recovered phylogeny and its associated support, it proved to be an important component of resolving relationships within the closely related Phocina. Changing which individuals were used to represent the species affected both the nuclear and mt topologies, sometimes with moderate bootstrap support. Using one *Halichoerus grypus* individual over the other for the nuclear dataset affected whether *Pusa* was recovered as monophyletic or not. For the mitochondrial dataset, regardless of which individuals were used, when only one individual/species was included, both *Phoca* species were drawn into the *Pusa-Halichoerus* clade, similar to the results recovered by Palo and Vainola (2006) in their investigation of the Phocina using mt genes. This underscores the importance of individual selection and use in shallow, rapid radiations. Based on this result, use of two individuals per species is recommended when investigating such clades to highlight areas of uncertainty and initially identify if there is any effect of the particular individual used, even if moderate clade support is recovered. If use of different individuals leads to the recovery of different topologies, using more individuals could be important in confidently resolving the relationships. Increasing the number of individuals can help to identify coalescence events and avoid artifactual resolution from incomplete lineage sorting or any 'non-phylogenetic' signal effects. In addition to being a promising technique for resolving the Phocina phylogeny, increasing both the number of individuals and loci may be important in unraveling the very close species relationships within Otariidae.

*Species monophyly*

One benefit of mtDNA over nuclear DNA at the taxonomic level investigated here was its ability to resolve species relationships. MtDNA strongly supported all species relationships, whether clade support was determined using BI, MLBP, or MPBP techniques. Conversely, while nuclear analysis generally recovered species as monophyletic, separating between closely related species was less clear. *Phoca largha* monophyly was not strongly supported with nuclear data, particularly when the locus *GHR* was removed. Though MP analysis of the NuclearMixed dataset recovered monophyly, MP analysis of nuclear DNA found *P. largha* to be paraphyletic with respect to *P. vitulina*. Individual *P. largha2* (Spotted seal ID:5879, Tables 5-2, 5-3) was recovered as the sister to both *P. vitulina* individuals. This result was not observed with other

analyses, though the second-most supported topology by BI of the NuclearMixed dataset recovered the same paraphyletic relationship (tree BPP=0.168). Two explanations are possible. First, different individuals were used for mt and nuclear analysis, as the mt genomes were downloaded from GenBank and most nuclear sequences were newly obtained (Table 5-3). It is possible that *P. largha2*, collected and identified as *P. largha* by field researchers, was, in fact, *P. vitulina*. This is unlikely, as experienced researchers in close contact with the animal performed the collection. The collection location is off the eastern coast of Kamchatka, Russia (Table 5-2), close to the northern edge of the *P. vitulina* range and *P. vitulina* are reported to be highly abundant in the nearby Commander Islands (Shaughnessy and Fay, 1977; King, 1983; Westlake and O'Corry-Crowe, 2002). Sympatry in this region is possible and the two species are very difficult to distinguish morphologically. They are separated based mainly on life history traits, particularly that *P. largha* are pagophilic and smaller, while *P. vitulina* are slightly larger and breed in ice-free rocky shores (Shaughnessy and Fay, 1977). A second, more probable hypothesis is that the two species are so closely related that there is not enough nuclear DNA information to consistently recover species monophyly. Historically, *P. largha* was considered a subspecies of *P. vitulina*, but ecological, geographical and some cranial differences led to their elevation to species-level status (for review, see Shaughnessy and Fay, 1977). *Phoca largha* and *P. vitulina* are known to interbreed occasionally (Shaughnessy and Fay, 1977) and this may contribute to the difficulty in recovering species monophyly using nuclear loci. It may also simply indicate that there has not been enough time for coalescence of some of the nuclear genes. Alternately, females have been shown to exhibit higher levels of breeding site fidelity than males (Harkonen and Harding, 2001), which could lead to faster isolation of mt lineages. While both misidentification and true genetic similarity are possible, given that *P. largha* monophyly is recovered most often, genetic similarity is preferred. Further study of both species across their ranges would be of interest, particularly in regions of sympatry.

**Taxonomic implications and conclusions**

The higher-level groupings, Pinnipedia, Otarioidea, and all subfamilies, tribes, and subtribes within Phocidae, were supported. Species relationships within Lobodontini and Phocina remain the only regions of the topology not confidently resolved. Within Lobodontini, two clades were recovered by both nuclear and combined mt+nuclear analyses, when the partitioning strategy selected by Bayes factors or the dataset excluding the partition illustrating

base compositional bias across taxa are removed. A sister relationship between *Ommatophoca rossii*, and *Lobodon carcinophagus* is proposed, although further study is required, given the generally low level of support. Within Phocina, there is good support for the monophyly of *Pusa*, consistent with morphology. Excluding mt 3rd positions to remove base composition bias across taxa and using two individuals per species appear to correct the problem that leads to the assumed artifactual recovery of *Halichoerus grypus* as sister to *Pusa caspica* found by many recent studies. Based solely on genetic data, *Halichoerus* is only differentiated from *Pusa* and *Phoca* at a species level, not genus level. The relationship between these three genera could not be definitively resolved. But irrespective of what the true resolution may be, Phocina was always very strongly supported and combining the three Phocina genera into one would yield a monophyletic genus from a phylogenetic perspective. If this is the case, *Pusa* Scopoli, 1777 has priority over *Halichoerus* (Nilsson, 1820), and *Phoca* has priority over both (Linnaeus, 1758). As proposed by Arnason and co-workers (Arnason et al., 1995), *Halichoerus grypus* could become *Phoca grypa*. However, with the present support for a monophyletic *Pusa*, retention of the three present genera is also acceptable from a phylogenetic perspective, despite their low genetic distances from one another. Given the morphological difference between *Halichoerus* and *Pusa* or *Phoca* and it would be premature to revoke its generic status until genetic relationships within Phocina are fully resolved and morphological review is performed.

**Table 5-1.** Taxonomy of the Phocidae (and otarioid outgroups). †Extinct. Not included in this study. *Only otariids included in this study are listed.

| Classification | Species name | Common Name | Authority |
|---|---|---|---|
| Order Carnivora | | | Bowdich, 1821 |
| Suborder Caniformia | | | Kretzoi, 1943 |
| Infraorder Arctoidea | | | Flower, 1869 |
| Pinnipedia | | | Illiger, 1811 |
| Superfamily Phocoidea | | | Smirnov 1908 |
| Family Phocidae (true, earless seals) | | | Gray, 1821 |
| Subfamily Phocinae (northern true seals) | | | Gray, 1821 |
| Tribe Erignathini | | | Chapskii 1955 |
| | *Erignathus barbatus* | Bearded seal | Erxleben, 1777 |
| Tribe Cystophorini | | | Burns & Fay, 1970 |
| | *Cystophora cristata* | Hooded seal | Erxleben, 1777 |
| Tribe Phocini | | | Gray, 1821 |
| Subtribe Histriophocina | | | Chapskii 1955 |
| | *HistrioHistriophoca fasciata* | Ribbon seal | Zimmerman, 1783 |
| | *Pagophilus groenlandicus* | Harp seal | Erxleben, 1777 |
| Subtribe Phocina | | | Chapskii 1955 |
| | *Phoca largha* | Spotted seal | Pallas, 1811 |
| | *Phoca vitulina* | Harbour seal | Linnaeus, 1758 |
| | *Halichoerus grypus* | Grey seal | Fabricius, 1791 |
| | *Pusa caspica* | Caspian seal | Gmelin, 1788 |
| | *Pusa sibirica* | Baikal seal | Gmelin, 1788 |
| | *Pusa hispida* | Ringed seal | Schreber, 1775 |
| Subfamily Monachinae ("southern" true seals) | | | Trouessart, 1897 |
| Tribe Monachini (monk seals) | | | Gray, 1869 |
| | *Monachus schauinslandi* | Hawaiian monk seal | Matschie, 1905 |
| | *Monachus monachus* | Mediterranean monk seal | Hermann, 1779 |
| | *Monachus tropicalis*† | Caribbean monk seal | Gray, 1850 |
| Tribe Mirounginí (elephant seals) | | | de Muizon, 1982 |
| | *Mirounga angustirostris* | Northern elephant seal | Gill, 1866 |
| | *Mirounga leonina* | Southern elephant seal | Linnaeus, 1758 |
| Tribe Lobodontini (antarctic seals) | | | Scheffer, 1958 |
| | *Lobodon carcinophagus* | Crabeater seal | Hombron & Jacquinot, 1842 |
| | *Ommatophoca rossii* | Ross seal | Gray, 1844 |
| | *Leptonychotes weddellii* | Weddell seal | Lesson, 1826 |
| | *Hydrurga leptonyx* | Leopard seal | Blainville, 1820 |
| Superfamily Otarioidea | | | Lucas, 1899 |
| Family Odobenidae (walrus) | | | Allen, 1880 |
| | *Odobenus rosmarus* | Walrus | Linnaeus, 1758 |
| Family Otariidae (fur seals & sea lions; eared seals)* | | | Gray, 1825 |
| Subfamily Otariinae / Arctocephalinae | | | Gray, 1825 / 1837 |
| | *Arctocephalus australis* | South American fur seal | Zimmerman, 1783 |
| | *Arctocephalus forsteri* | New Zealand fur seal | Lesson, 1828 |
| | *Eumetopias jubatus* | Steller's sea lion | Schreber, 1776 |

**Table 5-2.** Individual phocid identification and collection information. Parks Canada samples were obtained from the University of Alberta DNA repository location. DFO = Canadian Department of Fisheries and Oceans. APIS=Antarctic Pack-ice Seals program.

| Scientific Name | Individual ID | Collection ID | Sample Location | Received from: |
|---|---|---|---|---|
| *Cystophora cristata* | Hood 1626 | M20041626 | 'The Front', Newfoundland | Gary Stenson |
| | Hood 18384 | M18384 | 'The Front', Newfoundland | Gary Stenson |
| | Hood 1623 | M20041623 | 'The Front', Newfoundland | Gary Stenson |
| *Erignathus barbatus* | Beard SLI91 | 91 | 62 22'N -169 19'W (St. Lawrence Island, Alaska) | Bruce Robson |
| | Beard GA52 | 52 | 63 55' N 176 20'E (Gulf of Anadyr, Russia) | Bruce Robson |
| | Beard 1405 | M981405 | Labrador | Becky Sjare |
| *Pagophilus groenlandicus* | Harp 3511 | | Auyuttuq National Park, Canada | 11B-1f, Parks Canada |
| | Harp 1508 | M2006 1508 | 50 33.7N 54 25.5W (NAFO area 341, Labrador) | David McKinnon, DFO |
| | Harp 2144 | M2006 2144 | 50 15.0N 52 39.6W (NAFO area 344, Labrador) | David McKinnon, DFO |
| | Harp 46 | 96.46 | Front, Newfoundland | David McKinnon, DFO |
| *Halichoerus grypus* | Grey 19 | 96.19 | | 22I-5c, Parks Canada |
| | Grey 17 | 96.17 | | 22I-5a, Parks Canada |
| | Grey 18 | 96.18 | | 22I-5b, Parks Canada |
| *Pusa hispida* | Ringed 18 | Arv 18 | Arviat, Hudson Bay | J. Beauchesne |
| | Ringed 24 | HOL 94.24 | Holman, NWT | Lois Harwood |
| | Ringed 17 | Arv 17 | Arviat, Hudson Bay | J. Beauchesne |
| *Phoca vitulina concolor* | Harbour 88-6 | E88.6 | | 22I-5f, Parks Canada |
| | Harbour 88-5 | E88.5 | | 22I-5e, Parks Canada |
| *Phoca largha* | Spot 5879 | 5879 | 57.26628N 162.96752E (Ozernoy Gulf, Kamchatka, Russia) | Mike Cameron |
| | Spotted 04 | 06spotted04 | 59.451587N -177.935307W (Bering Sea) | Mike Cameron |
| | Spotted 03 | 06spotted03 | 58.550498N -173.922567W (Bering Sea) | Mike Cameron |
| *Pusa caspica* | Caspian 1 | #1 male | Caspian Sea | Bill Amos |
| | Caspian6 | #6 male | Caspian Sea | Bill Amos |
| | Caspian 7 | #7 female | Caspian Sea | Bill Amos |
| *Histriophoca fasciata* | Ribbon 5896 | 5896 male | 57.1992N 163.42879E (Ozernoy Gulf, Kamchatka, Russia) | Mike Cameron |
| | Ribbon 5898 | 5898 male | 57.29801N 163.0422E (Ozernoy Gulf, Kamchatka, Russia) | Mike Cameron |
| | Ribbon 5888 | 5888 female | 57.264953N 163.45768E (Ozernoy Gulf, Kamchatka, Russia) | Mike Cameron |
| | Ribbon 5876 | 5876 male | 57.29115N 163.0356E (Ozernoy Gulf, Kamchatka, Russia) | Mike Cameron |
| | Ribbon 05 | 06ribbon05 | 59.495295N -177.871812W (Bering Sea) | Mike Cameron |
| *Pusa sibirica* | Ps 11 | | Lake Baikal | Ulfur Arnason / Risto Vainola |
| *Ommatophoca rossii* | Ross 2506 | 2506 male | Ross Sea | APIS |
| | Ross 34 | R34 male | 69 30' S 01 23' W Queen Maud Land, S. Africa | A.S.Blix |
| | Ross 33 | R33 male | 69 28' S 01 18' W Queen Maud Land, S. Africa | A.S. Blix |
| | Ross 2507 | 2507 male | Ross Sea | APIS |

**Table 5-2.** *Continued.*

| Scientific Name | Individual ID | Collection ID | Sample Location | Received from: |
|---|---|---|---|---|
| *Hydrurga leptonyx* | Leop 3671 | B3671 | Bird Island | Ian Boyd / Tony Walker |
| | Leop 3675 | B3675 | Bird Island | Ian Boyd / Tony Walker |
| | Leop 2677 | L2677 male | Ross Sea | APIS |
| *Leptonychotes weddellii* | Weddell S9 | 8 male | 69.000S 39.583E Syowa Station | Katsufumi Sato |
| | Weddell S8 | 9 male | 69.000S 39.583E Syowa Station | Katsufumi Sato |
| | Weddell 9232 | 9232 | Big Razorback Island, Antarctica | Tom Gelatt |
| | Weddell 6624 | 6624 | Big Razorback Island, Antarctica | Tom Gelatt |
| *Mirounga angustirostris* | NES 2013 | ES2013, Drvot | San Mateo county, CA | Marine mammal center, CA |
| | NES 160 | Neseal 160 | California | Brent Stuart |
| | NES 142 | Neseal 142 | California | Brent Stuart |
| *Monachus schauinslandi* | HMS 2594 | 2594 male | Aukai, Laysan Island | Sara Iverson |
| | HMS 2072 | 2072 male | Aukai, Laysan Island | Sara Iverson |
| | HMS 2107 | 2107 female | Aukai, Laysan Island | Sara Iverson |
| | HMS 01 | | Aukai, Laysan Island | Shannon Atkinson |
| *Monachus monachus* | MMS | Dimitri | | Bill Amos |
| *Odobenus rosmarus* | Walrus 1 | 4148 | Northwest Territories, Canada | 11B-2c, Parks Canada |
| *Arctocephaus australis* | SAFS 3 | 03D male | Punta San Juan, Peru | Neil Gemmell |
| | SAFS 4 | 04D male | Punta San Juan, Peru | Neil Gemmell |
| *Arctocephalus forsteri* | NZFS 5 | OM5 male | Open Bay Island | Neil Gemmell |
| | NZFS 3 | OM3 male | Open Bay Island | Neil Gemmell |
| | NZFS 4 | OM4 male | Open Bay Island | Neil Gemmell |
| *Eumetopias jubatus* | Stellar 1b | Ej2 | Alaska, USA | 9G-1b, Parks Canada |
| *Mustela nivalis* | Least | | Alberta, Canada | Bob McClymont |
| *Meles meles* | Eurbadger | | | David Coltman |
| *Enhydra lutris* | Sea otter | | Prince William Sound, Alaska | Kim Kloecker, USGS |
| *Canis lupus* | Wolf 978-3 | | Banks Island, Northwest Territories | John Nagy |
| *Alopex lagopus* | MBAL 01 | | Split Lake, Manitoba | Dean Berezanski |
| *Felis catus* | Cat 22G-7b | | | 22G-7b, Parks Canada |
| *Lynx canadensis* | Lynx 5744 | | | 3E-6a, Parks Canada |

**Table 5-3.** GenBank accession numbers. Each box indicates either the individual used (new sequences, in bold) or the study reference (for previously published sequences) on top and the GenBank accession number on the bottom, where applicable. New sequences can be obtained from the author.

| Species | *IRBP* | *FES* | *GHR* | *CHRNA1* |
|---|---|---|---|---|
| *Cystophora cristata* | Fulton & Strobeck 2006<br>DQ205893 | Fulton & Strobeck 2006<br>DQ205782 | Fulton & Strobeck 2006<br>DQ205816 | Fulton & Strobeck 2006<br>DQ205742 |
| | **Hood 1626** | **Hood 1626** | **Hood 1626** | **Hood 1626** |
| *Erignathus barbatus* | Fulton & Strobeck 2006<br>DQ205894 | Fulton & Strobeck 2006<br>DQ205783 | Fulton & Strobeck 2006<br>DQ205817 | Fulton & Strobeck 2006<br>DQ205743 |
| | - | **Beard SLI91** | **Beard SLI91** | **Beard SLI91** |
| *Pagophilus groenlandicus* | Fulton & Strobeck 2006<br>DQ205901 | Fulton & Strobeck 2006<br>DQ205790 | Fulton & Strobeck 2006<br>DQ205825 | Fulton & Strobeck 2006<br>DQ205750 |
| | **Harp 3511** | **Harp 3511** | **Harp 3511** | **Harp 3511** |
| *Halichoerus grypus* | Fulton & Strobeck 2006<br>DQ205902 | Fulton & Strobeck 2006<br>DQ205791 | Fulton & Strobeck 2006<br>DQ205826 | Fulton & Strobeck 2006<br>DQ205751 |
| | **Grey 19** | **Grey 19** | **Grey 19** | **Grey 19** |
| *Pusa hispida* | Fulton & Strobeck 2006<br>DQ205899 | Fulton & Strobeck 2006<br>DQ205788 | Fulton & Strobeck 2006<br>DQ205823 | **Ringed 24** |
| | **Ringed 18** | **Ringed 18** | **Ringed 18** | **Ringed 18** |
| *Phoca vitulina* | **Harbour 88-6** | **Harbour 88-6** | **Harbour 88-6** | Fulton & Strobeck 2006<br>DQ205752 |
| | **Harbour 88-5** | **Harbour 88-5** | **Harbour 88-5** | **Harbour 88-5** |
| *Phoca largha* | Fulton & Strobeck 2006<br>DQ205904 | Fulton & Strobeck 2006<br>DQ205793 | Fulton & Strobeck 2006<br>DQ205827 | Fulton & Strobeck 2006<br>DQ205754 |
| | **Spot 5879** | **Spot 5879** | **Spot 5879** | **Spot 5879** |
| *Pusa caspica* | Fulton & Strobeck 2006<br>DQ205905 | **Caspian 1** | Fulton & Strobeck 2006<br>DQ205828 | Fulton & Strobeck 2006<br>DQ205755 |
| | **Caspian 6** | **Caspian 6** | **Caspian 6** | **Caspian 6** |
| *Histriophoca fasciata* | **Ribbon 5896** | **Ribbon 5896** | **Ribbon 5896** | **Ribbon 5896** |
| | **Ribbon 5898** | **Ribbon 5898** | **Ribbon 5898** | **Ribbon 5898** |
| *Pusa sibirica* | **Ps11** | **Ps11** | **Ps11** | **Ps11** |
| *Ommatophoca rossii* | Fulton & Strobeck 2006<br>DQ205900 | Fulton & Strobeck 2006<br>DQ205789 | Fulton & Strobeck 2006<br>DQ205824 | Fulton & Strobeck 2006<br>DQ205749 |
| | **Ross 2506** | **Ross 2506** | **Ross 2506** | **Ross 2506** |
| *Hydrurga leptonyx* | Fulton & Strobeck 2006<br>DQ205895 | - | Fulton & Strobeck 2006<br>DQ205818 | Fulton & Strobeck 2006<br>DQ205744 |
| | **Leop 3671** | **Leop 3671** | **Leop 3671** | **Leop 3671** |
| *Lobodon carcinophagus* | Fulton & Strobeck 2006<br>DQ205896 | Fulton & Strobeck 2006<br>DQ205784 | Fulton & Strobeck 2006<br>DQ205819 | Fulton & Strobeck 2006<br>DQ205745 |
| | **Ce 2123** | **Ce 2123** | **Ce 2123** | **Ce 2123** |
| *Leptonychotes weddellii* | Fulton & Strobeck 2006<br>DQ205903 | Fulton & Strobeck 2006<br>DQ205792 | Weddell 6624 | Fulton & Strobeck 2006<br>DQ205753 |
| | - | - | **Weddell S9** | **Weddell S9** |

**Table 5-3.** *Continued.*

| Species | *IRBP* | *FES* | *GHR* | *CHRNA1* |
|---|---|---|---|---|
| *Mirounga angustirostris* | Fulton & Strobeck 2006 DQ205897 | Fulton & Strobeck 2006 DQ205785 | Fulton & Strobeck 2006 DQ205820 | Fulton & Strobeck 2006 DQ205746 |
| | **NES 2013** | **NES 2013** | **NES 2013** | **NES 2013** |
| *Monachus schauinslandi* | Fulton & Strobeck 2006 DQ205898 | Fulton & Strobeck 2006 DQ205786 | Fulton & Strobeck 2006 DQ205821 | Fulton & Strobeck 2006 DQ205747 |
| | - | **HMS 2594** | **HMS 2594** | **HMS 2594** |
| *Monachus monachus* | **MMS** | Fulton & Strobeck 2006 DQ205787 | Fulton & Strobeck 2006 DQ205822 | Fulton & Strobeck 2006 DQ205748 |
| *Odobenus rosmarus* | Fulton & Strobeck 2006 DQ205892 | Fulton & Strobeck 2006 DQ205781 | Fulton & Strobeck 2006 DQ205815 | Fulton & Strobeck 2006 DQ205741 |
| *Arctocephalus australis* | Fulton & Strobeck 2006 DQ205887 | Fulton & Strobeck 2006 DQ205776 | Fulton & Strobeck 2006 DQ205810 | Fulton & Strobeck 2006 DQ205738 |
| *Arctocephalus forsteri* | Fulton & Strobeck 2006 DQ205888 | Fulton & Strobeck 2006 DQ205777 | Fulton & Strobeck 2006 DQ205811 | **NZFS5** |
| *Eumetopias jubatus* | Fulton & Strobeck 2006 DQ205890 | Fulton & Strobeck 2006 DQ205779 | Fulton & Strobeck 2006 DQ205813 | Fulton & Strobeck 2006 DQ205740 |
| *Mustela nivalis* | Sato et al. 2003 AB82973 | Fulton & Strobeck 2006 DQ205775 | Fulton & Strobeck 2006 DQ205809 | Fulton & Strobeck 2006 DQ205737 |
| *Meles meles* | Sato et al. 2003 AB082980 | Koepfli & Wayne 2003 AF498178 | Koepfli & Wayne 2003 AF498202 | Koepfli & Wayne 2003 AF498147 |
| *Enhydra lutris* | Sato et al. 2003 AB082978 | Koepfli & Wayne 2003 AF498162 | Koepfli & Wayne 2003 AF498186 | Koepfli & Wayne 2003 AF498131 |
| *Canis lupus* | Fulton & Strobeck 2006 DQ205907 | Fulton & Strobeck 2006 DQ205795 | Fulton & Strobeck 2006 DQ205836 | Fulton & Strobeck 2006 DQ205757 |
| *Alopex lagopus* | Fulton & Strobeck 2006 205908 | Fulton & Strobeck 2006 DQ205796 | Fulton & Strobeck 2006 DQ205837 | Fulton & Strobeck 2006 DQ205758 |
| *Felis catus* | Stanhope et al. 1992 Z11811 | - | Fulton & Strobeck 2006 DQ205829 | Fulton & Strobeck 2006 DQ205759 |
| *Lynx canadensis* | Fulton & Strobeck 2006 DQ205910 | Fulton & Strobeck 2006 DQ205797 | Fulton & Strobeck 2006 DQ205830 | Fulton & Strobeck 2006 DQ205760 |

| Species | *RHO* | *ADORA3* | *APOB* | *RAG1* | *RAG2* |
|---|---|---|---|---|---|
| *Cystophora cristata* | Fulton & Strobeck 2006 DQ205855 | **Hood 18384** | **Hood 1623** | **Hood 18384** | **Hood 18384** |
| | **Hood 1626** | **Hood 1626** | **Hood 1626** | **Hood 1626** | **Hood 1626** |
| *Erignathus barbatus* | Fulton & Strobeck 2006 DQ205856 | **Beard GA52** | **Beard GA52** | **Beard GA52** | **Beard GA52** |
| | **BeardSLI91** | **Beard 1405** | **Beard SLI91** | **Beard SLI91** | **Beard SLI91** |
| *Pagophilus groenlandicus* | Fulton & Strobeck 2006 DQ205863 | **Harp 1508** | **Harp 46** | **Harp 46** | **Harp 46** |
| | **Harp 3511** | **Harp 3511** | **Harp 3511** | **Harp 3511** | **Harp 3511** |
| *Halichoerus grypus* | Fulton & Strobeck 2006 DQ205864 | **Grey 17** | **Grey 17** | **Grey 17** | **Grey 17** |
| | **Grey 19** | **Grey 18** | **Grey 18** | **Grey 18** | **Grey 18** |
| *Pusa hispida* | Fulton & Strobeck 2006 DQ205864 | **Ringed 24** | **Ringed 17** | **Ringed 24** | **Ringed 24** |
| | **Ringed 18** | **Ringed 18** | **Ringed 18** | **Ringed 18** | **Ringed 18** |
| *Phoca vitulina* | Fulton & Strobeck 2006 DQ205865 | **Harbour 88-6** | **Harbour 88-6** | **Harbour 88-6** | **Harbour 88-6** |
| | **Harbour 88-5** | **Harbour 88-5** | **Harbour 88-5** | **Harbour 88-5** | **Harbour 88-5** |

**Table 5-3.** *Continued.*

| Species | *RHO* | *ADORA3* | *APOB* | *RAG1* | *RAG2* |
|---|---|---|---|---|---|
| *Phoca largha* | Fulton & Strobeck 2006 DQ205867 | Spotted 04 | Spotted 04 | Spotted 04 | Spotted 04 |
| | Spotted 5879 | Spotted 5879 | Spotted 5879 | Spotted 5879 | Spotted 5879 |
| *Pusa caspica* | Caspian 7 | Caspian 1 | Caspian 1 | Caspian 1 | Caspian 1 |
| | Caspian 6 | Caspian 6 | Caspian 6 | Caspian 6 | Caspian 6 |
| *Histriophoca fasciata* | Ribbon 5896 | Ribbon 5896 | Ribbon 5896 | Ribbon 5896 | Ribbon 5896 |
| | Ribbon 5898 | Ribbon 5898 | Ribbon 5898 | Ribbon 5898 | Ribbon 5898 |
| *Pusa sibirica* | Ps11 | Ps11 | Ps11 | Ps11 | Ps11 |
| *Ommatophoca rossii* | Ross 34 | Ross 34 | Ross 34 | Ross 34 | Ross 34 |
| | Ross 2507 | Ross 2507 | Ross 2507 | Ross 2507 | Ross 2506 |
| *Hydrurga leptonyx* | Fulton & Strobeck 2006 DQ205857 | Leop 2677 | Leop 2677 | Leop 3675 | Leop 3675 |
| | Leop 3671 | Leop 3671 | Leop 3671 | Leop 3671 | Leop 3671 |
| *Lobodon carcinophagus* | Fulton & Strobeck 2006 DQ205858 | Ce 6 | Ce 6 | Ce 2122 | Ce 2122 |
| | Ce 2123 | Ce 2123 | Ce 2122 | Ce 2123 | Ce 2123 |
| *Leptonychotes weddellii* | Fulton & Strobeck 2006 DQ205866 | Weddell 9232 | Weddell 6624 | Weddell 6624 | Weddell 6624 |
| | Weddell S9 | Weddell S9 | Weddell S8 | Weddell S9 | Weddell S9 |
| *Mirounga angustirostris* | Fulton & Strobeck 2006 DQ205859 | NES 160 | NES 160 | NES 160 | NES 160 |
| | NES 2013 | NES 2013 | NES 2013 | NES 2013 | NES 142 |
| *Monachus schauinslandi* | Fulton & Strobeck 2006 DQ205860 | HMS 2594 | HMS 2594 | HMS 2594 | HMS 2594 |
| | HMS 2594 | HMS 2072 | HMS01 | HMS2107 | HMS01 |
| *Monachus monachus* | Fulton & Strobeck 2006 DQ205861 | MMS | MMS | MMS | MMS |
| *Odobenus rosmarus* | Fulton & Strobeck 2006 DQ205854 | Walrus 1 | Walrus 1 | Walrus 1 | Walrus 1 |
| *Arctocephalus australis* | Fulton & Strobeck 2006 DQ205851 | SAFS 3 | SAFS 5 | SAFS 3 | SAFS 3 |
| *Arctocephalus forsteri* | NZFS5 | NZFS 5 | NZFS 5 | NZFS 3 | NZFS 3 |
| *Eumetopias jubatus* | Fulton & Strobeck 2006 DQ205853 | Stellar 1b | Stellar 1b | Stellar 1b | Stellar 1b |
| *Mustela nivalis* | Fulton & Strobeck 2006 DQ205850 | Least | Least | Least | Least |
| *Meles meles* | Koepfli & Wayne 2003 AF498226 | Eurbadger | Eurbadger | Eurbadger | Eurbadger |
| *Enhydra lutris* | Koepfli & Wayne 2003 AF498210 | Sea otter | Sato et al. 2006 AB193403 | Sea otter | Koepfli et al. 2007 DQ660280 |
| *Canis lupus* | Fulton & Strobeck 2006 DQ205869 | Wolf 9783 | Wolf 9783 | Wolf 9783 | - |
| *Alopex lagopus* | Fulton & Strobeck 2006 DQ205870 | MBA L01 | Arcfox | Arcfox | - |

**Table 5-3.** *Continued.*

| Species | *RHO* | *ADORA3* | *APOB* | *RAG1* | *RAG2* |
|---|---|---|---|---|---|
| *Felis catus* | Fulton & Strobeck 2006 DQ205871 | Cat 22G-7b | Cat 22G-7b | Cat 22G-7b | Johnson et al. 2006 Sci. DQ082330 |
| *Lynx canadensis* | Fulton & Strobeck 2006 DQ205872 | Lynx 5744 | Lynx 5744 | Lynx 5744 | Johnson et al. 2006 Sci. DQ082347 |

| Species | *BDNF* | *PNOC* | *APP* | *BRCA1, Fr.1* | *BRCA1, Fr.2* |
|---|---|---|---|---|---|
| *Cystophora cristata* | Hood 18384 | Hood 18384 | Hood 1623 | Hood 18384 | Hood 18384 |
| | Hood 1626 | Hood 1626 | Hood 1626 | Hood 1626 | Hood 1626 |
| *Erignathus barbatus* | Beard GA52 | Beard GA52 | Beard GA52 | Beard GA52 | Beard GA52 |
| | Beard SLI91 | Beard SLI91 | Beard SLI91 | Beard 1405 | Beard 1405 |
| *Pagophilus groenlandicus* | Harp 46 | Harp 1508 | Harp 1508 | Harp 46 | Harp 1508 |
| | Harp 3511 | Harp 3511 | Harp 3511 | Harp 3511 | Harp 3511 |
| *Halichoerus grypus* | Grey 17 | Grey 17 | Grey 17 | Grey 17 | Grey 17 |
| | Grey 18 | Grey 18 | Grey 19 | Grey 18 | Grey 18 |
| *Pusa hispida* | Ringed 24 | Ringed 24 | Ringed 24 | Ringed 24 | Ringed 24 |
| | Ringed 18 | Ringed 18 | Ringed 18 | Ringed 17 | Ringed 18 |
| *Phoca vitulina* | Harbour 88-6 | Harbour 88-6 | Harbour 88-6 | Harbour 88-6 | Harbour 88-6 |
| | Harbour 88-5 | Harbour 88-5 | Harbour 88-5 | Harbour 88-5 | Harbour 88-5 |
| *Phoca largha* | Spotted 04 | Spotted 04 | Spotted 04 | Spotted 04 | Spotted 04 |
| | Spotted 5879 | Spotted 5879 | Spotted 5879 | Spotted 5879 | Spotted 5879 |
| *Pusa caspica* | Caspian 1 | Caspian 1 | Caspian 1 | Caspian 1 | Caspian 1 |
| | Caspian 6 | Caspian 6 | Caspian 6 | Caspian 6 | Caspian 6 |
| *Histriophoca fasciata* | Ribbon 05 | Ribbon 05 | Ribbon 05 | Ribbon 05 | Ribbon 05 |
| | Ribbon 5898 | Ribbon 5896 | Ribbon 5896 | Ribbon 5896 | Ribbon 5898 |
| *Pusa sibirica* | Ps11 | Ps11 | Ps11 | Ps11 | Ps11 |
| *Ommatophoca rossii* | Ross 34 | Ross 34 | Ross 34 | - | Ross 34 |
| | Ross 2506 | Ross 2506 | Ross 2506 | - | Ross 2506 |
| *Hydrurga leptonyx* | Leop 2677 | Leop 3675 | Leop 2677 | Leop 2677 | Leop 2677 |
| | Leop 3671 | Leop 3671 | Leop 3671 | Leop 3671 | Leop 3671 |
| *Lobodon carcinophagus* | Ce 6 | Ce 2122 | Ce 2122 | Ce 2122 | Ce 2122 |
| | Ce 2123 | Ce 2123 | Ce 2123 | Ce 2123 | Ce 2123 |

**Table 5-3.** *Continued.*

| Species | *BDNF* | *PNOC* | *APP* | *BRCA1, Fr.1* | *BRCA1, Fr.2* |
|---|---|---|---|---|---|
| *Leptonychotes weddellii* | Weddell 6624 | Weddell 6624 | Weddell 6624 | Weddell 6624 | Weddell 6624 |
| | Weddell S9 | Weddell S9 | Weddell S8 | Weddell S9 | Weddell S9 |
| *Mirounga angustirostris* | NES 160 | NES 160 | NES 160 | NES 160 | NES 160 |
| | NES 2013 | NES 2013 | NES 2013 | NES 2013 | NES 2013 |
| *Monachus schauinslandi* | HMS 2594 | HMS 2594 | HMS 2594 | HMS2107 | HMS2107 |
| | HMS01 | HMS01 | HMS01 | HMS 2072 | HMS 2072 |
| *Monachus monachus* | MMS | MMS | MMS | MMS | MMS |
| *Odobenus rosmarus* | Walrus 1 | Walrus 1 | Walrus 1 | Pollinger et al. 2005 DQ240423 | Walrus 1 |
| *Arctocephalus australis* | SAFS 4 | SAFS 4 | SAFS 3 | SAFS 4 | SAFS 4 |
| *Arctocephalus forsteri* | NZFS 5 | NZFS 4 | NZFS 3 | NZFS 5 | NZFS 5 |
| *Eumetopias jubatus* | Stellar 1b | Stellar 1b | Stellar 1b | Stellar 1b | Stellar 1b |
| *Mustela nivalis* | Least | Least | Least | Least | Least |
| *Meles meles* | Eurbadger | Eurbadger | Eurbadger | Eurbadger | Eurbadger |
| *Enhydra lutris* | Sea otter | Koepfli et al. 2007 DQ660253 | Sea otter | Sea otter | Sea otter |
| *Canis lupus* | Wolf 9783 | Wolf 9783 | Wolf 9783 | Pollinger et al. 2005 DQ240429 | Pollinger et al. 2005 DQ240429 |
| *Alopex lagopus* | MBA L01 | MBA L01 | MBA L01 | Pollinger et al. 2005 DQ240424 | Pollinger et al. 2005 DQ240424 |
| *Felis catus* | Cat 22G-7b | Johnson et al. 2006 DQ082285 | Johnson et al. 2006 DQ081730 | Madsen et al. 2001 AF284018 | Madsen et al. 2001 AF284018 |
| *Lynx canadensis* | Lynx 5744 | Johnson et al. 2006 DQ082302 | Lynx 5744 | Lynx 5744 | - |

| Species | *CREM* | *FLVCR1* | *PLCB4* | *COI* |
|---|---|---|---|---|
| *Cystophora cristata* | Hood 18384 | Hood 18384 | Hood 18384 | Davis et al. 2004 AY377144 |
| | Hood 1626 | Hood 1626 | Hood 1626 | Arnason et al. 2006 AM181028 |
| *Erignathus barbatus* | Beard GA52 | Beard GA52 | Beard GA52 | Davis et al. 2004 AY377143 |
| | Beard 1405 | Beard 1405 | Beard 1405 | Arnason et al. 2006 AM181027 |
| *Pagophilus groenlandicus* | Harp 1508 | Harp 1508 | Harp 1508 | Davis et al. 2004 AY377145 |
| | Harp 3511 | Harp 3511 | Harp 2144 | Arnason et al. 2006 AM181030 |
| *Halichoerus grypus* | Grey 17 | Grey 17 | Grey 17 | Arnason et al. 1993 NC_001602 |
| | Grey 18 | Grey 18 | Grey 18 | Grey 18 |
| *Pusa hispida* | Ringed 17 | Ringed 24 | Ringed 24 | Davis et al. 2004 AY377146 |
| | Ringed 18 | Ringed 18 | Ringed 18 | Arnason et al. 2006 AM181036 |

**Table 5-3.** *Continued.*

| Species | *CREM* | *FLVCR1* | *PLCB4* | *COI* |
|---|---|---|---|---|
| *Phoca vitulina* | Harbour 88-6 | Harbour 88-6 | Harbour 88-6 | Arnason & Johnsson 1992 NC_001325 |
| | Harbour 88-5 | Harbour 88-5 | Harbour 88-5 | Arnason et al. 2006 AM181032 |
| *Phoca largha* | Spotted 04 | Spotted 04 | Spotted 04 | Davis et al. 2004 AY377147 |
| | Spotted 5879 | Spotted 5879 | Spotted 5879 | Arnason et al. 2006 AM181031 |
| *Pusa caspica* | Caspian 1 | Caspian 1 | Caspian 1 | Caspian 1 |
| | Caspian 6 | Caspian 6 | Caspian 6 | Arnason et al. 2006 AM181033 |
| *Histriophoca fasciata* | Spotted 5896 | Spotted 04 | Spotted 5896 | Ribbon 5888 |
| | Spotted 5898 | Spotted 5898 | Spotted 5898 | Arnason et al. 2006 AM181029 |
| *Pusa sibirica* | Ps11 | Ps11 | Ps11 | Arnason et al. 2006 AM181034 |
| | - | - | - | Arnason et al. 2006 AM181035 |
| *Ommatophoca rossii* | Ross 34 | Ross 33 | Ross 33 | Davis et al. 2004 AY377132 |
| | Ross 2506 | Ross 2507 | Ross 2506 | Ross 2506 |
| *Hydrurga leptonyx* | Leop 2677 | Leop 2677 | Leop 3675 | Davis et al. 2004 AY377134 |
| | Leop 3671 | Leop 3671 | Leop 3671 | Arnason et al. 2006 AM181026 |
| *Lobodon carcinophagus* | Ce 2122 | Ce 2122 | Ce 2122 | Davis et al. 2004 AY377130 |
| | Ce 2123 | Ce 2123 | Ce 2123 | Arnason et al. 2006 AM181024 |
| *Leptonychotes weddellii* | Weddell 6624 | Weddell 6624 | Weddell 6624 | Davis et al. 2004 AY377136 |
| | Weddell S9 | Weddell S9 | Weddell S9 | Arnason et al. 2006 AM181025 |
| *Mirounga angustirostris* | NES 160 | NES 160 | NES 160 | Davis et al. 2004 AY377138 |
| | NES 2013 | NES 2013 | NES 2013 | - |
| *Mirounga leonina* | - | - | - | Davis et al. 2004 AY377140 |
| | - | - | - | Arnason et al. 2006 AM181023 |
| *Monachus schauinslandi* | HMS 2594 | HMS 2594 | HMS 2594 | Davis et al. 2004 AY377141 |
| | HMS 01 | HMS 01 | HMS 2107 | Arnason et al. 2006 AM181022 |
| *Monachus monachus* | MMS | MMS | MMS | Davis et al. 2004 AY377142 |
| *Odobenus rosmarus* | Walrus 1 | Walrus 1 | Walrus 1 | Davis et al. 2004 AY377148 |
| | - | - | - | Arnason et al. 2002 AJ428576 |
| *Arctocephalus australis* | SAFS 4 | SAFS 3,4,5 | SAFS 3 | Davis et al. 2004 AY377150 |
| *Arctocephalus forsteri* | NZFS4 | NZFS 3,4 | NZFS 3 | Lin et al., 2002 AF513820 |

**Table 5-3.** *Continued.*

| Species | *CREM* | *FLVCR1* | *PLCB4* | *COI* |
|---|---|---|---|---|
| *Eumetopias jubatus* | **Stellar 1b** | **Stellar 1b** | **Stellar 1b** | Arnason et al., 2002 AJ428578 |
| *Mustela nivalis* | **Least** | **Least** | **Least** | Fulton & Strobeck 2007 DQ533939 |
| *Meles meles* | **Eurbadger** | **Eurbadger** | **Eurbadger** | **Eurbadger** |
| *Enhydra lutris* | **Sea otter** | **Sea otter** | **Sea otter** | Fulton & Strobeck 2007 DQ533938 |
| *Canis lupus* | **Wolf 9783** | **Wolf 9783** | **Wolf 9783** | Delisle & Strobeck 2005 AY598496 |
| *Alopex lagopus* | **MBA L01** | **MBA L01** | **MBA L01** | Delisle & Strobeck 2005 AY598508 |
| *Felis catus* | **Cat 22G-7b** | **Cat 22G-7b** | **Cat 22G-7b** | Lopez et al. 1996 NC_001700 |
| *Lynx canadensis* | **Lynx 5744** | **Lynx 5744** | **Lynx 5744** | Delisle & Strobeck 2005 AY598472 |

| Species | *COII* | *COIII* | *CYTB* | *ND1* |
|---|---|---|---|---|
| *Cystophora cristata* | Davis et al. 2004 AY377167 | Davis et al. 2004 AY377259 | Palo & Vainola 2006 AY140981 | Davis et al. 2004 AY377357 |
| | Arnason et al. 2006 AM181028 | Arnason et al. 2006 AM181028 | Arnason et al. 2006 AM181028 | Arnason et al. 2006 AM181028 |
| *Erignathus barbatus* | Davis et al. 2004 AY377166 | Davis et al. 2004 AY377258 | Palo & Vainola 2006 AY140982 | Davis et al. 2004 AY377356 |
| | Arnason et al. 2006 AM181027 | Arnason et al. 2006 AM181027 | Arnason et al. 2006 AM181027 | Arnason et al. 2006 AM181027 |
| *Pagophilus groenlandicus* | Davis et al. 2004 AY377168 | Davis et al. 2004 AY377260 | Harp96.46 | Davis et al. 2004 AY377358 |
| | Arnason et al. 2006 AM181030 | Arnason et al. 2006 AM181030 | Arnason et al. 2006 AM181030 | Arnason et al. 2006 AM181030 |
| *Halichoerus grypus* | Arnason et al. 1993 NC_001602 | Arnason et al. 1993 NC_001602 | Arnason et al. 1993 NC_001602 | Arnason et al. 1993 NC_001602 |
| | **Grey 18** | **Grey 18** | **Grey 18** | **Grey 18** |
| *Pusa hispida* | Davis et al. 2004 AY377169 | Davis et al. 2004 AY377261 | Palo & Vainola 2006 AY140976 | Davis et al. 2004 AY377359 |
| | Arnason et al. 2006 AM181036 | Arnason et al. 2006 AM181036 | Arnason et al. 2006 AM181036 | Arnason et al. 2006 AM181036 |
| *Phoca vitulina* | Arnason & Johnsson 1992 NC_001325 | Arnason & Johnsson 1992 NC_001325 | Arnason & Johnsson 1992 NC_001325 | Arnason & Johnsson 1992 NC_001325 |
| | Arnason et al. 2006 AM181032 | Arnason et al. 2006 AM181032 | Arnason et al. 2006 AM181032 | Arnason et al. 2006 AM181032 |
| *Phoca largha* | Davis et al. 2004 AY377170 | Davis et al. 2004 AY377262 | Palo & Vainola 2006 AY140979 | Davis et al. 2004 AY377360 |
| | Arnason et al. 2006 AM181031 | Arnason et al. 2006 AM181031 | Arnason et al. 2006 AM181031 | Arnason et al. 2006 AM181031 |
| *Pusa caspica* | **Caspian 1** | **Caspian 1** | **Caspian 1** | **Caspian 1** |
| | Arnason et al. 2006 AM181033 | Arnason et al. 2006 AM181033 | Arnason et al. 2006 AM181033 | Arnason et al. 2006 AM181033 |
| *Histriophoca fasciata* | **Ribbon 5888** | **Ribbon 5876** | **Ribbon 5888** | **Ribbon 5876** |
| | Arnason et al. 2006 AM181029 | Arnason et al. 2006 AM181029 | Arnason et al. 2006 AM181029 | Arnason et al. 2006 AM181029 |
| *Pusa sibirica* | Arnason et al. 2006 AM181034 | Arnason et al. 2006 AM181034 | Arnason et al. 2006 AM181034 | Arnason et al. 2006 AM181034 |

**Table 5-3.** *Continued.*

| Species | *COII* | *COIII* | *CYTB* | *NDI* |
|---|---|---|---|---|
| *P. sibirica con't.* | Arnason et al. 2006 AM181035 | Arnason et al. 2006 AM181035 | Arnason et al. 2006 AM181035 | Arnason et al. 2006 AM181035 |
| *Ommatophoca rossii* | Davis et al. 2004 AY377155 | Davis et al. 2004 AY377247 | Davis et al. 2004 AY377322 | Davis et al. 2004 AY377349 |
| | **Ross 2506** | **Ross 2506** | **Ross 2506** | **Ross 2506** |
| *Hydrurga leptonyx* | Davis et al. 2004 AY377157 | Davis et al. 2004 AY377249 | Davis et al. 2004 AY377323 | Davis et al. 2004 AY377350 |
| | Arnason et al. 2006 AM181026 | Arnason et al. 2006 AM181026 | Arnason et al. 2006 AM181026 | Arnason et al. 2006 AM181026 |
| *Lobodon carcinophagus* | Davis et al. 2004 AY377153 | Davis et al. 2004 AY377245 | Davis et al. 2004 AY377321 | Davis et al. 2004 AY377348 |
| | Arnason et al. 2006 AM181024 | Arnason et al. 2006 AM181024 | Arnason et al. 2006 AM181024 | Arnason et al. 2006 AM181024 |
| *Leptonychotes weddellii* | Davis et al. 2004 AY377159 | Davis et al. 2004 AY377251 | Davis et al. 2004 AY377324 | Davis et al. 2004 AY377351 |
| | Arnason et al. 2006 AM181025 | Arnason et al. 2006 AM181025 | Arnason et al. 2006 AM181025 | Arnason et al. 2006 AM181025 |
| *Mirounga angustirostris* | Davis et al. 2004 AY377161 | Davis et al. 2004 AY377253 | Davis et al. 2004 AY377325 | Davis et al. 2004 AY377352 |
| *Mirounga leonina* | Davis et al. 2004 AY377163 | Davis et al. 2004 AY377255 | Davis et al. 2004 AY377326 | Davis et al. 2004 AY377353 |
| | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 |
| *Monachus schauinslandi* | Davis et al. 2004 AY377164 | Davis et al. 2004 AY377256 | **HMS 01** | Davis et al. 2004 AY377354 |
| | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 |
| *Monachus monachus* | Davis et al. 2004 AY377165 | Davis et al. 2004 AY377257 | Davis et al. 2004 AY378327 | Davis et al. 2004 AY377355 |
| *Odobenus rosmarus* | Davis et al. 2004 AY377171 | Davis et al. 2004 AY377263 | **Walrus1** | Davis et al. 2004 AY377361 |
| | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 |
| *Arctocephalus australis* | Davis et al. 2004 AY377173 | Davis et al. 2004 AY377265 | Davis et al. 2004 AY377329 | Davis et al. 2004 AY377363 |
| *Arctocephalus forsteri* | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 |
| *Eumetopias jubatus* | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 |
| *Mustela nivalis* | **Least** | **Least** | Fulton & Strobeck 2007 DQ533943 | **Least** |
| *Meles meles* | **Eurbadger** | **Eurbadger** | Ledje & Arnason 1996a X94922 | **Eurbadger** |
| *Enhydra lutris* | **Sea otter** | **Sea otter** | Fulton & Strobeck 2007 DQ533942 | **Sea otter** |
| *Canis lupus* | Delisle & Strobeck 2005 AY598497 | Delisle & Strobeck 2005 AY598498 | Delisle & Strobeck 2005 AY598499 | Delisle & Strobeck 2005 AY598500 |
| *Alopex lagopus* | Delisle & Strobeck 2005 AY598509 | Delisle & Strobeck 2005 AY598510 | Delisle & Strobeck 2005 AY598511 | Delisle & Strobeck 2005 AY598512 |
| *Felis catus* | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 |
| *Lynx canadensis* | Delisle & Strobeck 2005 AY598473 | Delisle & Strobeck 2005 AY598474 | Delisle & Strobeck 2005 AY598475 | Delisle & Strobeck 2005 AY598476 |

**Table 5-3.** *Continued.*

| Species | *ND2* | *ND3* | *ND4* | *ND4L* |
|---|---|---|---|---|
| *Cystophora cristata* | Davis et al. 2004<br>AY377277 | Davis et al. 2004<br>AY377213 | Davis et al. 2004<br>AY377339 | Davis et al. 2004<br>AY377236 |
| | Arnason et al. 2006<br>AM181028 | Arnason et al. 2006<br>AM181028 | Arnason et al. 2006<br>AM181028 | Arnason et al. 2006<br>AM181028 |
| *Erignathus barbatus* | Davis et al. 2004<br>AY377276 | Davis et al. 2004<br>AY377212 | Davis et al. 2004<br>AY377338 | Davis et al. 2004<br>AY377235 |
| | Arnason et al. 2006<br>AM181027 | Arnason et al. 2006<br>AM181027 | Arnason et al. 2006<br>AM181027 | Arnason et al. 2006<br>AM181027 |
| *Pagophilus groenlandicus* | Davis et al. 2004<br>AY377278 | Davis et al. 2004<br>AY377214 | Davis et al. 2004<br>AY377340 | Davis et al. 2004<br>AY377237 |
| | Arnason et al. 2006<br>AM181030 | Arnason et al. 2006<br>AM181030 | Arnason et al. 2006<br>AM181030 | Arnason et al. 2006<br>AM181030 |
| *Halichoerus grypus* | Arnason et al. 1993<br>NC_001602 | Arnason et al. 1993<br>NC_001602 | Arnason et al. 1993<br>NC_001602 | Arnason et al. 1993<br>NC_001602 |
| | **Grey 18** | **Grey 18** | **Grey 18** | **Grey 18** |
| *Pusa hispida* | Davis et al. 2004<br>AY377279 | Davis et al. 2004<br>AY377215 | Davis et al. 2004<br>AY377341 | Davis et al. 2004<br>AY377238 |
| | Arnason et al. 2006<br>AM181036 | Arnason et al. 2006<br>AM181036 | Arnason et al. 2006<br>AM181036 | Arnason et al. 2006<br>AM181036 |
| *Phoca vitulina* | Arnason & Johnsson 1992<br>NC_001325 | Arnason & Johnsson 1992<br>NC_001325 | Arnason & Johnsson 1992<br>NC_001325 | Arnason & Johnsson 1992<br>NC_001325 |
| | Arnason et al. 2006<br>AM181032 | Arnason et al. 2006<br>AM181032 | Arnason et al. 2006<br>AM181032 | Arnason et al. 2006<br>AM181032 |
| *Phoca largha* | Davis et al. 2004<br>AY377280 | Davis et al. 2004<br>AY377216 | Davis et al. 2004<br>AY377342 | Davis et al. 2004<br>AY377239 |
| | Arnason et al. 2006<br>AM181031 | Arnason et al. 2006<br>AM181031 | Arnason et al. 2006<br>AM181031 | Arnason et al. 2006<br>AM181031 |
| *Pusa caspica* | **Caspian 1** | **Caspian 1** | **Caspian 1** | **Caspian 1** |
| | Arnason et al. 2006<br>AM181033 | Arnason et al. 2006<br>AM181033 | Arnason et al. 2006<br>AM181033 | Arnason et al. 2006<br>AM181033 |
| *Histriophoca fasciata* | **Ribbon 5888** | **Ribbon 5888** | **Ribbon 5876** | **Ribbon 5888** |
| | Arnason et al. 2006<br>AM181029 | Arnason et al. 2006<br>AM181029 | Arnason et al. 2006<br>AM181029 | Arnason et al. 2006<br>AM181029 |
| *Pusa sibirica* | Arnason et al. 2006<br>AM181034 | Arnason et al. 2006<br>AM181034 | Arnason et al. 2006<br>AM181034 | Arnason et al. 2006<br>AM181034 |
| | Arnason et al. 2006<br>AM181035 | Arnason et al. 2006<br>AM181035 | Arnason et al. 2006<br>AM181035 | Arnason et al. 2006<br>AM181035 |
| *Ommatophoca rossii* | Davis et al. 2004<br>AY377269 | Davis et al. 2004<br>AY377201 | Davis et al. 2004<br>AY377331 | Davis et al. 2004<br>AY377224 |
| | **Ross 2506** | **Ross 2506** | **Ross 2506** | **Ross 2506** |
| *Hydrurga leptonyx* | Davis et al. 2004<br>AY377270 | Davis et al. 2004<br>AY377203 | Davis et al. 2004<br>AY377332 | Davis et al. 2004<br>AY377226 |
| | Arnason et al. 2006<br>AM181026 | Arnason et al. 2006<br>AM181026 | Arnason et al. 2006<br>AM181026 | Arnason et al. 2006<br>AM181026 |
| *Lobodon carcinophagus* | Davis et al. 2004<br>AY377268 | Davis et al. 2004<br>AY377199 | Davis et al. 2004<br>AY377330 | Davis et al. 2004<br>AY377222 |
| | Arnason et al. 2006<br>AM181024 | Arnason et al. 2006<br>AM181024 | Arnason et al. 2006<br>AM181024 | Arnason et al. 2006<br>AM181024 |
| *Leptonychotes weddellii* | Davis et al. 2004<br>AY377271 | Davis et al. 2004<br>AY377205 | Davis et al. 2004<br>AY377333 | Davis et al. 2004<br>AY377228 |
| | Arnason et al. 2006<br>AM181025 | Arnason et al. 2006<br>AM181025 | Arnason et al. 2006<br>AM181025 | Arnason et al. 2006<br>AM181025 |

**Table 5-3.** *Continued.*

| Species | *ND2* | *ND3* | *ND4* | *ND4L* |
|---|---|---|---|---|
| *Mirounga angustirostris* | Davis et al. 2004 AY377272 | Davis et al. 2004 AY377207 | Davis et al. 2004 AY377334 | Davis et al. 2004 AY377230 |
| *Mirounga leonina* | Davis et al. 2004 AY377273 | Davis et al. 2004 AY377209 | Davis et al. 2004 AY377335 | Davis et al. 2004 AY377232 |
| | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 |
| *Monachus schauinslandi* | Davis et al. 2004 AY377274 | Davis et al. 2004 AY377210 | Davis et al. 2004 AY377336 | Davis et al. 2004 AY377233 |
| | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 |
| *Monachus monachus* | Davis et al. 2004 AY377275 | Davis et al. 2004 AY377211 | Davis et al. 2004 AY377337 | Davis et al. 2004 AY377234 |
| *Odobenus rosmarus* | Davis et al. 2004 AY377281 | Davis et al. 2004 AY377217 | Davis et al. 2004 AY377343 | Davis et al. 2004 AY377240 |
| | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 |
| *Arctocephalus australis* | Davis et al. 2004 AY377283 | Davis et al. 2004 AY377219 | Davis et al. 2004 AY377345 | Davis et al. 2004 AY377242 |
| *Arctocephalus forsteri* | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 |
| *Eumetopias jubatus* | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 |
| *Mustela nivalis* | Fulton & Strobeck 2007 DQ533949 | **Least** | **Least** | **Least** |
| *Meles meles* | **Eurbadger** | **Eurbadger** | **Eurbadger** | **Eurbadger** |
| *Enhydra lutris* | Fulton & Strobeck 2007 DQ533948 | **Sea otter** | **Sea otter** | **Sea otter** |
| *Canis lupus* | Delisle & Strobeck 2005 AY598501 | Delisle & Strobeck 2005 AY598502 | Delisle & Strobeck 2005 AY598503 | Delisle & Strobeck 2005 AY598504 |
| *Alopex lagopus* | Delisle & Strobeck 2005 AY598513 | Delisle & Strobeck 2005 AY598514 | Delisle & Strobeck 2005 AY598515 | Delisle & Strobeck 2005 AY598516 |
| *Felis catus* | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 |
| *Lynx canadensis* | Delisle & Strobeck 2005 AY598477 | Delisle & Strobeck 2005 AY598478 | Delisle & Strobeck 2005 AY598479 | Delisle & Strobeck 2005 AY598480 |

| Species | *ND5* | *ATP6* | *ATP8* | *12S rRNA* |
|---|---|---|---|---|
| *Cystophora cristata* | Davis et al. 2004 AY377376 | Davis et al. 2004 AY377306 | Davis et al. 2004 AY377190 | **Hood 18384** |
| | Arnason et al. 2006 AM181028 | Arnason et al. 2006 AM181028 | Arnason et al. 2006 AM181028 | Arnason et al. 2006 AM181028 |
| *Erignathus barbatus* | Davis et al. 2004 AY377375 | Davis et al. 2004 AY377305 | Davis et al. 2004 AY377189 | **Beard SLI91** |
| | Arnason et al. 2006 AM181027 | Arnason et al. 2006 AM181027 | Arnason et al. 2006 AM181027 | **Beard 1405** |
| *Pagophilus groenlandicus* | Davis et al. 2004 AY377377 | Davis et al. 2004 AY377307 | Davis et al. 2004 AY377191 | **Harp 46** |
| | Arnason et al. 2006 AM181030 | Arnason et al. 2006 AM181030 | Arnason et al. 2006 AM181030 | Arnason et al. 2006 AM181030 |
| *Halichoerus grypus* | Arnason et al. 1993 NC_001602 | Arnason et al. 1993 NC_001602 | Arnason et al. 1993 NC_001602 | Arnason et al. 1993 NC_001602 |
| | **Grey 18** | **Grey 18** | **Grey 18** | **Grey 18** |
| *Pusa hispida* | Davis et al. 2004 AY377378 | Davis et al. 2004 AY377308 | Davis et al. 2004 AY377192 | **Ringed 24** |
| | Arnason et al. 2006 AM181036 | Arnason et al. 2006 AM181036 | Arnason et al. 2006 AM181036 | Arnason et al. 2006 AM181036 |

**Table 5-3.** *Continued.*

| Species | *ND5* | *ATP6* | *ATP8* | *12S rRNA* |
|---|---|---|---|---|
| *Phoca vitulina* | Arnason & Johnsson 1992 NC_001325 | Arnason & Johnsson 1992 NC_001325 | Arnason & Johnsson 1992 NC_001325 | Arnason & Johnsson 1992 NC_001325 |
| | Arnason et al. 2006 AM181032 | Arnason et al. 2006 AM181032 | Arnason et al. 2006 AM181032 | Arnason et al. 2006 AM181032 |
| *Phoca largha* | Davis et al. 2004 AY377379 | Davis et al. 2004 AY377309 | Davis et al. 2004 AY377193 | **Spotted03** |
| | Arnason et al. 2006 AM181031 | Arnason et al. 2006 AM181031 | Arnason et al. 2006 AM181031 | Arnason et al. 2006 AM181031 |
| *Pusa caspica* | **Caspian 1** | **Caspian 1** | **Caspian 1** | **Caspian 1** |
| | Arnason et al. 2006 AM181033 | Arnason et al. 2006 AM181033 | Arnason et al. 2006 AM181033 | Arnason et al. 2006 AM181033 |
| *Histriophoca fasciata* | **Ribbon 5876** | **Ribbon 5888** | **Ribbon 5888** | **Ribbon 5888** |
| | Arnason et al. 2006 AM181029 | Arnason et al. 2006 AM181029 | Arnason et al. 2006 AM181029 | Arnason et al. 2006 AM181029 |
| *Pusa sibirica* | Arnason et al. 2006 AM181034 | Arnason et al. 2006 AM181034 | Arnason et al. 2006 AM181034 | Arnason et al. 2006 AM181034 |
| | Arnason et al. 2006 AM181035 | Arnason et al. 2006 AM181035 | Arnason et al. 2006 AM181035 | Arnason et al. 2006 AM181035 |
| *Ommatophoca rossii* | Davis et al. 2004 AY377368 | Davis et al. 2004 AY377294 | Davis et al. 2004 AY377178 | Davis et al. 2004 AY377287 |
| | **Ross 2506** | **Ross 2506** | **Ross 2506** | **Ross 2506** |
| *Hydrurga leptonyx* | Davis et al. 2004 AY377369 | Davis et al. 2004 AY377296 | Davis et al. 2004 AY377180 | Davis et al. 2004 AY377288 |
| | Arnason et al. 2006 AM181026 | Arnason et al. 2006 AM181026 | Arnason et al. 2006 AM181026 | Arnason et al. 2006 AM181026 |
| *Lobodon carcinophagus* | Davis et al. 2004 AY377367 | Davis et al. 2004 AY377292 | Davis et al. 2004 AY377176 | Davis et al. 2004 AY377286 |
| | Arnason et al. 2006 AM181024 | Arnason et al. 2006 AM181024 | Arnason et al. 2006 AM181024 | Arnason et al. 2006 AM181024 |
| *Leptonychotes weddellii* | Davis et al. 2004 AY377370 | Davis et al. 2004 AY377298 | Davis et al. 2004 AY377182 | Davis et al. 2004 AY377289 |
| | Arnason et al. 2006 AM181025 | Arnason et al. 2006 AM181025 | Arnason et al. 2006 AM181025 | Arnason et al. 2006 AM181025 |
| *Mirounga angustirostris* | Davis et al. 2004 AY377371 | Davis et al. 2004 AY377303 | Davis et al. 2004 AY377184 | Davis et al. 2004 AY377290 |
| | - | - | - | **NES 160** |
| *Mirounga leonina* | Davis et al. 2004 AY377372 | Davis et al. 2004 AY377302 | Davis et al. 2004 AY377186 | Davis et al. 2004 AY377291 |
| | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 | Arnason et al. 2006 AM181023 |
| *Monachus schauinslandi* | Davis et al. 2004 AY377373 | Davis et al. 2004 AY377303 | Davis et al. 2004 AY377187 | **HMS 01** |
| | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 | Arnason et al. 2006 AM181022 |
| *Monachus monachus* | Davis et al. 2004 AY377374 | Davis et al. 2004 AY377304 | Davis et al. 2004 AY377188 | **MMS** |
| *Odobenus rosmarus* | Davis et al. 2004 AY377380 | Davis et al. 2004 AY377310 | Davis et al. 2004 AY377194 | **Walrus 1** |
| | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 | Arnason et al. 2002 AJ428576 |
| *Arctocephalus australis* | Davis et al. 2004 AY377382 | Davis et al. 2004 AY377312 | Davis et al. 2004 AY377196 | **SAFS 3** |

**Table 5-3.** *Continued.*

| Species | *ND5* | *ATP6* | *ATP8* | *12S rRNA* |
|---|---|---|---|---|
| *Arctocephalus forsteri* | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 | Lin et al., 2002 AF513820 |
| *Eumetopias jubatus* | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 | Arnason et al., 2002 AJ428578 |
| *Mustela nivalis* | **Least** | **Least** | **Least** | Ledje & Arnason 1996b Y08515 |
| *Meles meles* | **Eurbadger** | **Eurbadger** | **Eurbadger** | Ledje & Arnason 1996b Y08513 |
| *Enhydra lutris* | **Sea otter** | **Sea otter** | **Sea otter** | Ledje & Arnason 1996b Y08512 |
| *Canis lupus* | Delisle & Strobeck 2005 AY598505 | Delisle & Strobeck 2005 AY598494 | Delisle & Strobeck 2005 AY598495 | **Wolf** |
| *Alopex lagopus* | Delisle & Strobeck 2005 AY598517 | Delisle & Strobeck 2005 AY598506 | Delisle & Strobeck 2005 AY598507 | **Arctic fox** |
| *Felis catus* | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 | Lopez et al. 1996 NC_001700 | Ledje & Arnason 1996b Y08503 |
| *Lynx canadensis* | Delisle & Strobeck 2005 AY598481 | Delisle & Strobeck 2005 AY598470 | Delisle & Strobeck 2005 AY598471 | **Lynx** |

**Table 5-4.** Loci included in this study. **BRCA1* fragments 1 and 2 were concatenated for all analyses.

| Gene | Acronym | Fragment | Primer Reference | Forward Primer / Reverse Primer |
|---|---|---|---|---|
| Adenosine A3 receptor | *ADORA3* | exon 2 | Murphy et al., 2001 | ACCCCCATGTTTGGCTGGAA |
| | | | | GATAGGGTTCATCATGGAGTT |
| Apolipoprotein B | *APOB* | exon 26 | Jiang et al., 1998 (J1R) | GTGCCAGGTTCAATCAGTATAAGT |
| | | | Amrine-Madsen et al., 2003 (187F) | CCAGCAAAATTTTCTTTTACTTCAA |
| Brain-derived neurotrophic factor | *BDNF* | exon 1 | Murphy et al., 2001 | CATCCTTTTCCTTACTATGGTT |
| | | | | TTCCAGTGCCTTTTGTCTATG |
| Breast cancer 1, early onset* | *BRCA1* | exon 9 | Lindblad-Toh et al., 2005 | GCC ATG TGG CAC ARA TRC TC |
| (fragment 1) | *(BRCA1.1)* | (fragment 1) | | CTC TRC TTT CTT GAT AAA RTC CTC AG |
| Breast cancer 1, early onset* | *BRCA1* | exon 9 | Lindblad-Toh et al., 2005 | TCA AAG CGY CAG TCA TTT GC |
| (fragment 2) | *(BRCA1.2)* | (fragment 2) | | AAW CAG ACA TGG AGA GAA RTC |
| Interphotoreceptor binding protein | *IRBP* | exon 1 | Stanhope et al., 1992 | ATGGCCAAGGTCCTCTTGGATAACTACTGCTT |
| | | | (+217,-1531) | CGCAGGTCCATGATGAGGTGCTCCGTGTCCTG |
| | | | Fulton & Strobeck 2006 | GGTACAGTGCCGACAAAGATG |
| | | | (internal +785,-913) | GCTTCTGGAGGTCCAGGGC |
| Prepronociceptin | *PNOC* | exon 2 | Murphy et al., 2001 | GCATCCTTGAGTGTGAAGAGAA |
| | | | | TGCCTCATAAACTCACTGAACC |
| Recombination activating gene 1 | *RAG1* | exon 1 | Teeling et al., 2000 | GCTTTGATGGACATGGAAGAAGACAT |
| | | | (F1705,R2864) | GAGCCATCCCTATCAATAATTTCAGG |
| | | | Sato et al. 2004 | AGCCTCCCAAAATCTTGTCTTCCACTCCA |
| | | | (internal F2357,R2486) | AATGTCACAGTGAAGGGCATCTATGGAAGG |
| Recombination activating gene 2 | *RAG2* | exon 1 | Murphy et al., 2001 | TCATGGAGGGAAAACACCAAA |
| | | | | TGCACTGGAGACAGAGATTC |
| Cholinergic receptor, nicotinic, alpha 1 | *CHRNA1* | intron 8 | Lyons et al., 1997 | GACCATGAAGTCAGACCAGGAG |
| | | | | GGAGTATGTGGTCCATCACCAT |

**Table 5-4.** *Continued.*

| Gene | Acronym | Fragment | Primer Reference | Forward Primer / Reverse Primer |
|---|---|---|---|---|
| Feline sarcoma oncogene | *FES* | intron 14 | Venta et al., 1996 | GGGGAACTTTGGCGAAGTGTT |
| | | | | TCCATGACGATGTAGATGGG |
| Feline leukemia virus subgroup C cellular receptor 1 | *FLVCR1* | intron 7 | Housley et al., 2006 | TTGAAATCACTTACCCTGAATCTGA |
| | | | | TCCTTGAGCCAATGTGAACA |
| Growth hormone receptor | *GHR* | intron 9 | Venta et al., 1996 | CCAGTTCCAGTTCCAAAGAT |
| | | | | TGATTCTTCTGGTCAAGGCA |
| Rhodopsin | *RHO* | intron 3 | Venta et al., 1996 | TACATGTTCGTGGTCCACTT |
| | | | | TGGTGGGTGAAGATGTAGAA |
| Amyloid beta (A4) precursor protein | *APP* | 3' UTR | Murphy et al., 2001 | TCCAAGATGCAGCAGAACG |
| | | | | CTAATGTGTGCACATAAAACAGG |
| cAMP responsive element modulator | *CREM* | 3' UTR | Murphy et al., 2001 | AGGAACTCAAGGCCCTCAAA |
| | | | | GGGAGGACAAATGTCTTTCAA |
| Phospholipase C, beta 4 | *PLCB4* | 3' UTR | Murphy et al., 2001 | GTGAAATTGGAAGCCGAGAT |
| | | | | CACCAAGCTCATTTACTTGTGA |
| 12 mitochondrial protein-coding genes + 12S rRNA | mtDNA | mitochondrion | Delisle & Strobeck 2002 | |

**Table 5-5.** Informative insertion-deletion information. Indels are named by gene, indel type, and length. I=insertion, D=deletion, ID=undetermined. [1]Canids have missing information for this indel; it may define Caniformia.

| Group | # Indels | Information |
|---|---|---|
| Caniformia | 2 | *GHR* ID-5, ID-3, |
| total=14 | 2 | *CHRNA1* ID-18, ID-14 |
| | 1 | *RHO* ID-6 |
| | 1 | *BRCA1* ID-3 |
| | 1 | *PNOC* ID-3 |
| | 3 | *APP* ID-1 (x3) |
| | 1 | *PLCB4* ID-1 |
| | 1 | *FLVCR1* ID-7 |
| Canidae | 2 | *GHR* I-1, D-17 |
| total=14 | 4 | *CHRNA1* I-4, D-5, D-1 (x2) |
| | 1 | *RHO* I-1 |
| | 2 | *APP* I-1, I-3 |
| | 2 | *PLCB4* I-1 (x2) |
| | 3 | *FLVCR1* D-27, I-4, I-1 |
| Arctoidea | 1 | *FES* I-6, I-16 |
| total=6 | 2 | *GHR* D-1 |
| | 1 | *FLVCR1* I-9[1] |
| | 2 | *FLVCR1* D-1 (x2) |
| Mustelidae | 1 | *FES* D-1 |
| total=9 | 2 | *GHR* D-25, D-2 |
| | 1 | *CHRNA1* D-2 |
| | 1 | *BRCA1* D-3 |
| | 1 | *APP* D-3 |
| | 1 | *APOB* D-15 |
| | 2 | *PLCB4* D-3, D-11 |

| Group | # Indels | Information |
|---|---|---|
| Pinnipedia | 1 | *FES* D-30 |
| total=2 | 1 | *PLCB4* I-2 |
| Otarioidea | 1 | *FES* D-4 |
| total=4 | 1 | *GHR* D-12 |
| | 1 | *CHRNA1* I-6 |
| | 1 | *BRCA1* D-3 |
| Otariidae | 1 | *PLCB4* I-1 |
| Phocidae | 1 | *APP* D-1 |
| Phocinae | 1 | *GHR* D-7 |
| Monachinae | 1 | *GHR* D-4 |
| total=2 | 1 | *PLCB4* D-1 |
| **Species:** | | |
| *M. angustirostris* | 1 | *FES* D-1 |
| total=3 | 1 | *BRCA1* D-3 |
| | 1 | *APP* I-1 |
| *P. vitulina* | 1 | *BRCA1* D-9 |
| *E. barbatus* | 1 | *APP* D-1 (x2) |
| **Homoplasies:** | | |
| *A. lagopus* + *F. catus* | 1 | *BRCA1* D-3 |
| *C. lupus* + Mustelidae | 1 | *FLVCR1* D-4 |

**Table 5-6.** Model information and maximum parsimony (MP) results. PI=Parsimony informative, CI=Consistency index, RI=retention index.

| | DNA: | | | | | | | | | Amino acid: | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene | Model | Aligned length (bp) | #PI sites | Prop. PI sites | CI | RI | #MP trees | MP score | ILD p-value | Model | Aligned length (aa) | #PI sites | CI | RI | #MP trees | MP score |
| *ADORA3* | HKY+G | 354 | 58 | 0.215 | 0.813 | 0.890 | 6 | 107 | 0.89 | JTT+G+F | 117 | 20 | 0.816 | 0.91 | 1 | 38 |
| *APOB* | HKY+G | 940 | 155 | 0.214 | 0.899 | 0.940 | max=200 | 258 | 0.02* | JTT+G+F | 313 | 62 | 0.924 | 0.947 | max=200 | 119 |
| *BDNF* | HKY+I | 566 | 58 | 0.124 | 0.862 | 0.916 | 2 | 94 | 0.82 | JTT+I | 188 | 8 | 0.882 | 0.882 | 87 | 17 |
| *BRCA1* | GTR+G | 1416 | 242 | 0.265 | 0.856 | 0.903 | max=200 | 487 | 0.06 | JTT+G | 471 | 150 | 0.885 | 0.919 | max=200 | 321 |
| *IRBP* | HKY+I+G | 1177 | 200 | 0.243 | 0.816 | 0.870 | max=200 | 434 | 0.03* | JTT+G+F | 392 | 52 | 0.901 | 0.933 | 54 | 121 |
| *PNOC* | HKY+G | 268 | 37 | 0.224 | 0.906 | 0.940 | 24 | 85 | 0.43 | JTT | 89 | 15 | 0.946 | 0.957 | 14 | 37 |
| *RAG1* | SYM+I+G | 1108 | 112 | 0.172 | 0.803 | 0.862 | max=200 | 299 | 0.02* | JTT+G | 368 | 21 | 0.93 | 0.949 | max=200 | 43 |
| *RAG2* | HKY | 444 | 40 | 0.144 | 0.952 | 0.968 | 2 | 83 | 1.00 | JTT+I | 147 | 15 | 0.933 | 0.96 | 44 | 30 |
| *CHRNA1* | K80+G | 365 | 104 | 0.378 | 0.873 | 0.900 | 45 | 197 | 0.05* | - | - | - | - | - | - | - |
| *FES* | GTR+I | 415 | 116 | 0.347 | 0.878 | 0.903 | max=200 | 197 | 0.36 | - | - | - | - | - | - | - |
| *FLVCR1* | HKY+G | 278 | 83 | 0.345 | 0.936 | 0.967 | 16 | 125 | 0.19 | - | - | - | - | - | - | - |
| *GHR* | HKY+G | 633 | 133 | 0.302 | 0.901 | 0.935 | max=200 | 262 | 0.01** | - | - | - | - | - | - | - |
| *RHO* | HKY+G | 273 | 49 | 0.245 | 0.833 | 0.897 | 20 | 96 | 1.00 | - | - | - | - | - | - | - |
| *APP* | GTR+I | 650 | 81 | 0.163 | 0.917 | 0.953 | 18 | 132 | 0.83 | - | - | - | - | - | - | - |
| *CREM* | HKY+I | 358 | 25 | 0.101 | 0.950 | 0.956 | 90 | 40 | 0.81 | - | - | - | - | - | - | - |
| *PLCB4* | GTR+G | 323 | 74 | 0.266 | 0.876 | 0.940 | 18 | 121 | 0.79 | - | - | - | - | - | - | - |
| All Nuclear | | 9568 | 1567 | 0.164 | 0.837 | 0.903 | 3 | 2796 | - | - | - | - | - | - | - | - |
| ATP6 | GTR+I+G | 680 | 309 | 0.454 | 0.356 | 0.678 | 4 | 1431 | 0.49 | | | | | | | |
| ATP8 | HKY+I+G | 161 | 96 | 0.596 | 0.414 | 0.732 | 3 | 440 | 0.85 | | | | | | | |
| COI | GTR+I+G | 1545 | 579 | 0.375 | 0.334 | 0.691 | 1 | 2857 | 0.48 | | | | | | | |
| COII | HKY+I+G | 684 | 268 | 0.392 | 0.344 | 0.689 | 2 | 1211 | 0.74 | | | | | | | |
| COIII | HKY+I+G | 786 | 301 | 0.383 | 0.353 | 0.687 | 2 | 1440 | 0.27 | | | | | | | |
| CYTB | GTR+I+G | 1140 | 499 | 0.438 | 0.371 | 0.655 | 1 | 2149 | 0.32 | | | | | | | |
| ND1 | GTR+I+G | 957 | 384 | 0.401 | 0.385 | 0.662 | 26 | 1613 | 0.95 | | | | | | | |
| ND2 | GTR+I+G | 1044 | 512 | 0.490 | 0.432 | 0.703 | 8 | 2111 | 0.98 | | | | | | | |
| ND3 | GTR+I+G | 345 | 162 | 0.470 | 0.396 | 0.682 | 94 | 690 | 0.01** | | | | | | | |
| ND4 | GTR+I+G | 1377 | 640 | 0.465 | 0.372 | 0.673 | 1 | 2864 | 0.55 | | | | | | | |
| ND4L | HKY+I+G | 290 | 131 | 0.452 | 0.369 | 0.679 | 2 | 569 | 0.50 | | | | | | | |
| ND5 | GTR+I+G | 1827 | 874 | 0.478 | 0.391 | 0.688 | 1 | 3832 | 0.03* | | | | | | | |
| 12S rRNA | GTR+I+G | 961 | 253 | 0.263 | 0.484 | 0.751 | 1 | 899 | 0.89 | | | | | | | |
| mtDNA | | 11794 | 5008 | 0.425 | 0.377 | 0.683 | 1 | 29264 | - | | | | | | | |
| All Data | | 21362 | 6472 | 0.303 | 0.428 | 0.692 | 5 | 25035 | Mt vs. nuclear: 0.84 | | | | | | | |

*AllData does not include a second individual of walrus so the values are not a direct sum of the partitions.

**Table 5-7.** Clade support values for the Phocidae nuclear phylogeny. Node numbers refer to Figure 5-3. [1]*P. largha* paraphyletic (48%)

| | | NuclearMixed: | | | NuclearDNA: | | |
|---|---|---|---|---|---|---|---|
| | | by gene | | one partition | by gene | | one partition |
| Node | Description | BPP | ML BP | MP BP | BPP | ML BP | MP BP |
| 1 | Phocinae | 1.0 | 100 | 99 | 1.0 | 98 | 98 |
| 2 | Phocini | 1.0 | 99 | 97 | 0.99 | 63 | 81 |
| 3 | Histriophocina | 1.0 | 100 | 96 | 1.0 | 98 | 98 |
| 4 | *Pagophilus* | 1.0 | 89 | 78 | 1.0 | 99 | 98 |
| 5 | Phocina | 1.0 | 93 | 86 | 1.0 | 98 | 96 |
| 6 | *Phoca vitulina* | 1.0 | 100 | 98 | 1.0 | 100 | 98 |
| 7 | *Phoca largha* | 0.83 | 59 | 45 | 0.6 | 50 | N/A[1] |
| 8 | *Pusa+Halichoerus* | 1.0 | 78 | 74 | 1.0 | 70 | 67 |
| 9 | *Halichoerus* | 1.0 | 98 | 97 | 1.0 | 87 | 89 |
| 10 | *Pusa* | 1.0 | 63 | 64 | 1.0 | 60 | 58 |
| 11 | *P.sibirica+P.caspica* | 1.0 | 48 | 36 | 0.95 | 40 | 30 |
| 12 | *Pusa hispida* | 1.0 | 92 | 82 | 1.0 | 92 | 83 |
| 13 | Monachini | 1.0 | 95 | 95 | 1.0 | 96 | 96 |
| 14 | Miroungini+Lobodontini | 1.0 | 95 | 93 | 1.0 | 100 | 95 |
| 15 | Lobodontini | 1.0 | 96 | 75 | 1.0 | 100 | 95 |
| 16 | *Ommatophoca+Lobodon* | 1.0 | 57 | 67 | 0.95 | 61 | 66 |
| 17 | *Hydrurga leptonyx* | 1.0 | 96 | 99 | 1.0 | 100 | 100 |

**Table 5-8.** Clade support values for Phocidae mitochondrial phylogeny. **ML analysis with 3rd positions excluded recovers *Phoca+Pusa* (MLBP=82%); *Halichoerus-Phoca-Pusa* (MLBP=100%). Node numbers refer to Figure 5-4, letters indicate nodes recovered by alternate topologies (Figure 5-4 b-d, f-g).

| | | MtProt - by gene | | MtDNA - one partition | | | MtDNA partitioned | | | | No 3rd Positions | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Node | Description | BPP | ML BP | BPP | MLBP | MP BP | BPP by gene | BPP by pos | BPP by rate | ML BP by rate | LogDet ME BP | ML BP |
| 1 | Phocini | 1.0 | 98 | 1.0 | 100 | 100 | 1.0 | 1.0 | 1.0 | 100 | 100 | 100 |
| 2 | Histriophocina | 1.0 | 90 | 1.0 | 100 | 100 | 1.0 | 1.0 | 1.0 | 100 | 99 | 100 |
| 3 | *Pusa+Halichoerus* | 0.97 | 49 | 1.0 | 89 | 79 | 1.0 | 1.0 | 1.0 | 93 | 89 | ** |
| 4 | *P.sibirica, P.caspica, Halichoerus* | - | 39 | 0.52 | - | 5-4 (c) | - | - | 0.90 | 44 | 5-4 (b) | ** |
| 5 | *P.caspica+Halichoerus* | 0.95 | 56 | 0.94 | 62 | 5-4 (c) | 0.93 | 0.69 | 0.97 | 70 | 5-4 (b) | ** |
| c,d | *P. sibirica+P.hispida* | 0.60 | - | - | 38 | 37 | 0.7 | 0.6 | - | - | - | - |
| 6 | *Phoca* | 1.0 | 96 | 1.0 | 100 | 100 | 1.0 | 1.0 | 1.0 | 100 | 100 | 100 |
| 7 | Monachini | 1.0 | 98 | 1.0 | 100 | 99 | 1.0 | 1.0 | 1.0 | 100 | 97 | 99 |
| 8 | Miroungini+Lobodontini | 1.0 | 85 | 1.0 | 100 | 98 | 1.0 | 1.0 | 1.0 | 100 | 83 | 100 |
| 9 | Lobodontini | 1.0 | 96 | 1.0 | 100 | 100 | 1.0 | 1.0 | 1.0 | 100 | 100 | 97 |
| 10 | *O. rossii, L. weddellii, H. leptonyx* | 5-4 (g) | 5-4 (g) | 0.61 | 58 | 5-4 (g) | polytomy | 5-4 (g) | 0.93 | 60 | 5-4 (f) | 5-4 (f) |
| g | *L. carcinophagus, L. weddellii, H. leptonyx* | 1.0 | 67 | - | - | 62 | - | 0.97 | - | - | - | - |
| 11 | *Leptonychotes+Hydrurga* | 1.0 | 91 | 1.0 | 100 | 100 | 1.0 | 1.0 | 1.0 | 100 | 100 | 100 |

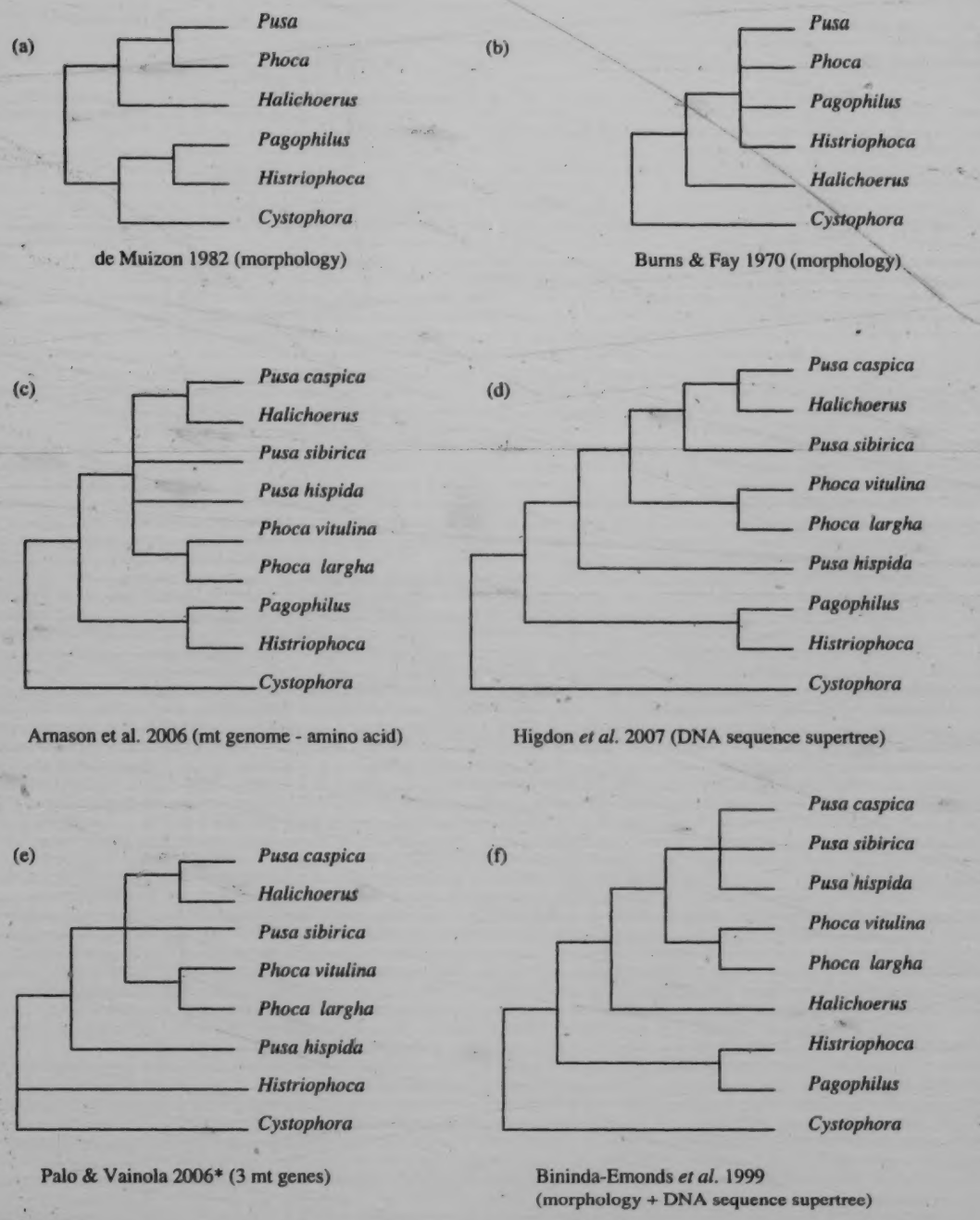


**Figure 5-1.** Review of proposed phocine relationships. (a,b) Examples of morphological studies of the Phocini (+ Cystophorini); (c-f) Recent molecular hypotheses that include all *Pusa* species. Nomenclature used is that of this study (for consistency) and does not represent the taxonomy used in the original studies c, e, and f. *Adapted – the basal polytomy indicates discrepancy between analysis methods, not a lack of resolution indicated by the authors.

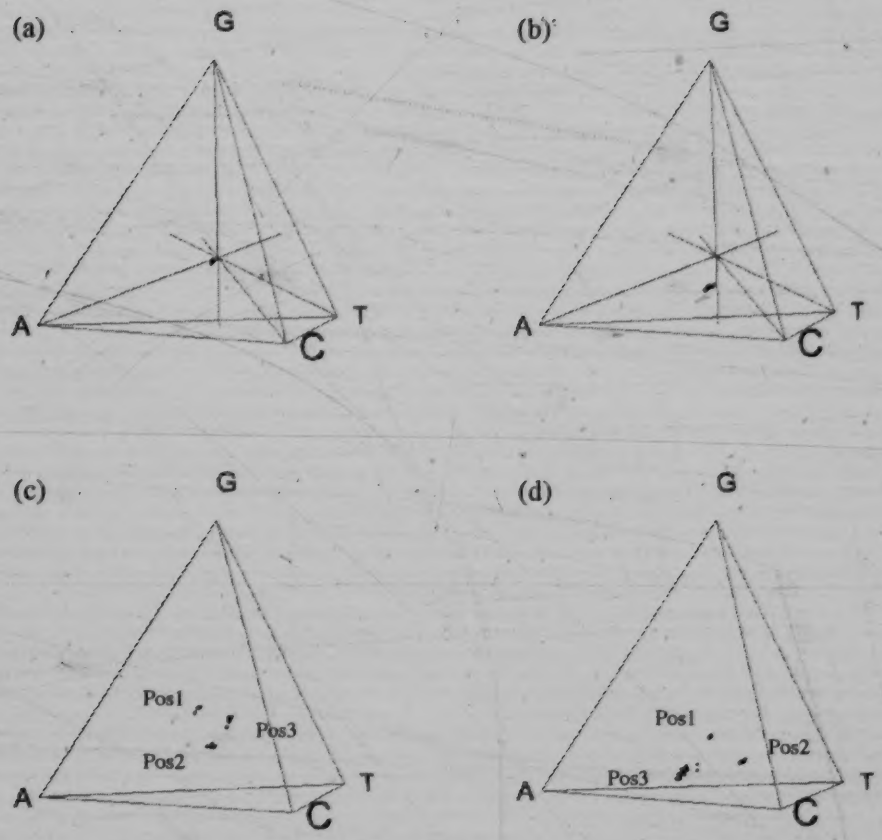


**Figure 5-2.** Visualization of nuclear and mitochondrial base composition. Each dot represents a taxon in the nuclear (a,c) or mitochondrial (b,d) matrix. The higher the proportion of a particular nucleotide (A,G,C,T), the closer the data point is placed to that corner of the tetrahedron. Spread of the data points indicates variation in composition across taxa. Matrices are visualized as either the entire combined matrix (a,b) or as only the protein-coding genes divided by codon position (c,d).

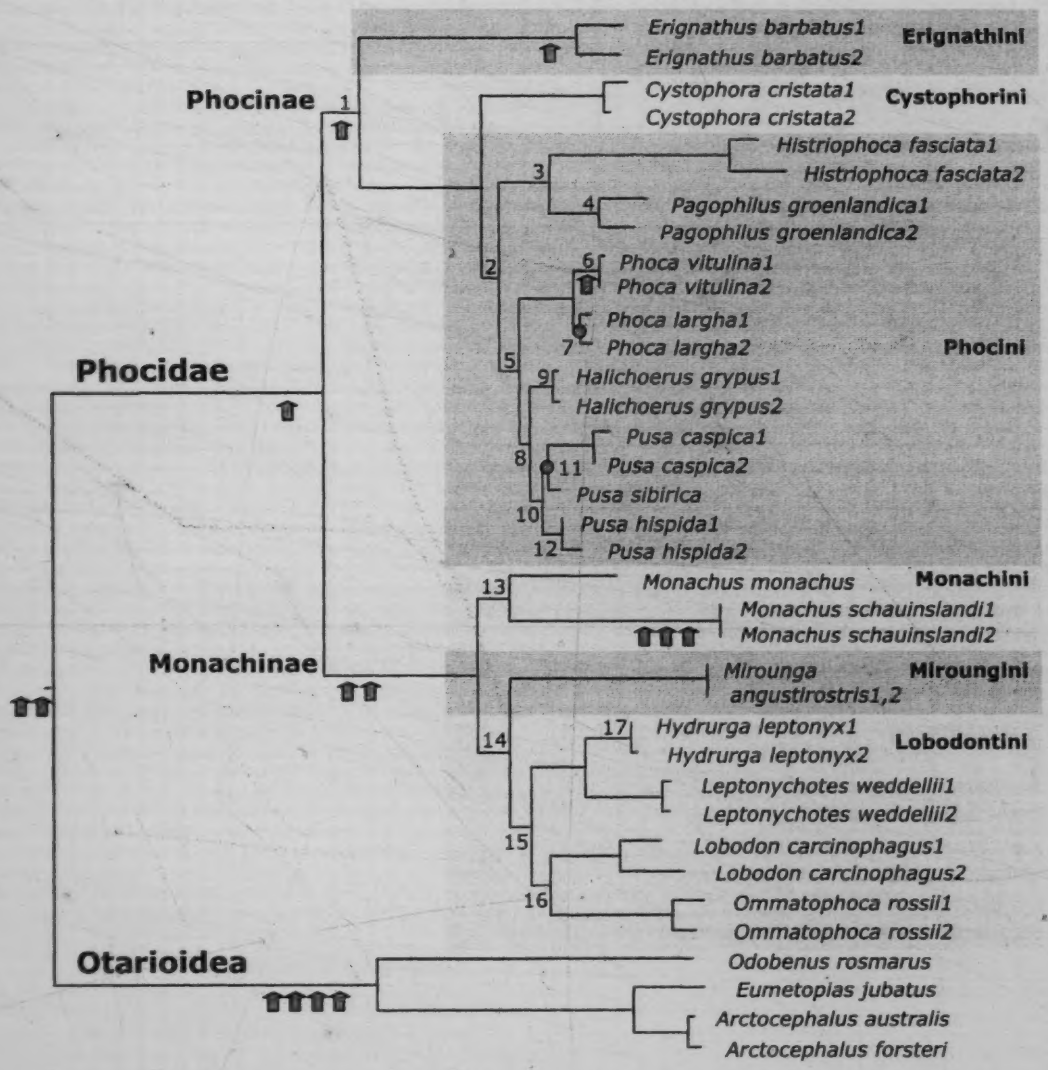


**Figure 5-3.** Nuclear phylogeny of the Phocidae. Branch lengths were optimized in RAxML for the nuclear dataset partitioned by gene, with the outgroups (mustelids, canids, felids) removed. All nuclear analyses yielded the same topology with highly similar relative branch lengths. All nodes were supported by Bayesian posterior probability=1.0 or ML or MP bootstrap ≥99%, except nodes that are numbered; support values are in Table 5-7. Circles represent nodes that were not supported by BP>50% or BPP>0.9 in all analyses. Arrows represent insertion/deletion events (Table 5-5).

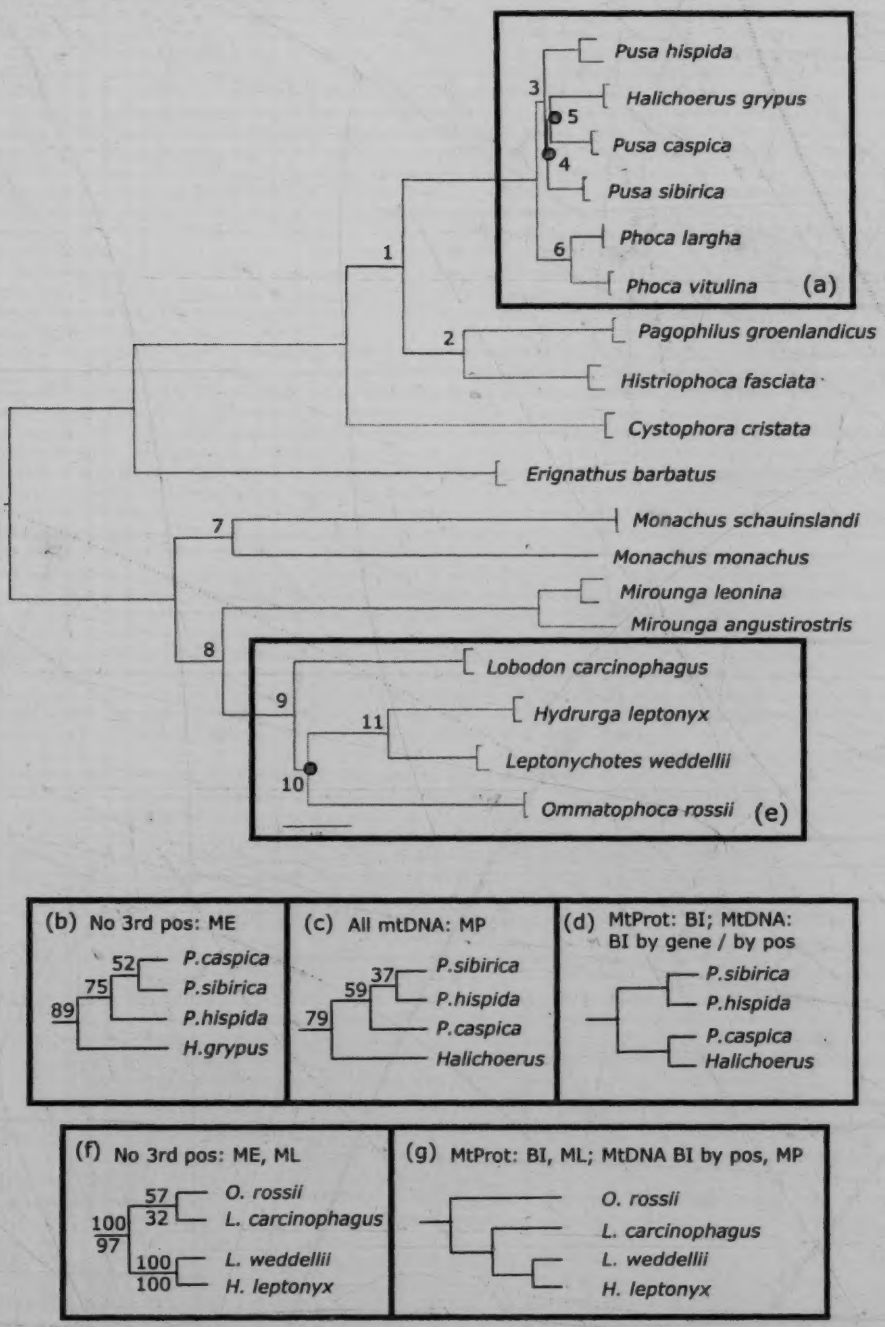


**Figure 5-4.** Mitochondrial phylogeny of the Phocidae. Branch lengths were estimated in RAxML, with the dataset partitioned by rate (see methods); outgroups were removed afterwards. (a-d) Different topologies recovered for Phocina by different methods. (e-g) Different topologies recovered for the Lobodontini by different methods. Nodes that were not supported by BPP=1.0 or BP=100% in all methods are numbered and their support values are listed in Table 5-8. Circles represent clades that were not supported by BPP>0.90 or BP>50% in all analyses.

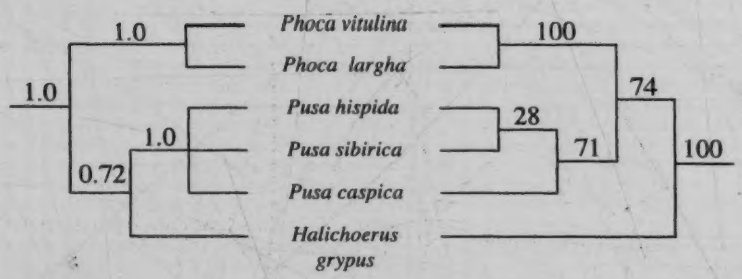


**Figure 5-5 (a).** Phocina topology recovered by combined nuclear (no *GHR*) + mitochondrial (no 3rd positions) dataset. Left: Bayesian posterior clade probability (BPP). Right: maximum likelihood bootstrap (MLBP).

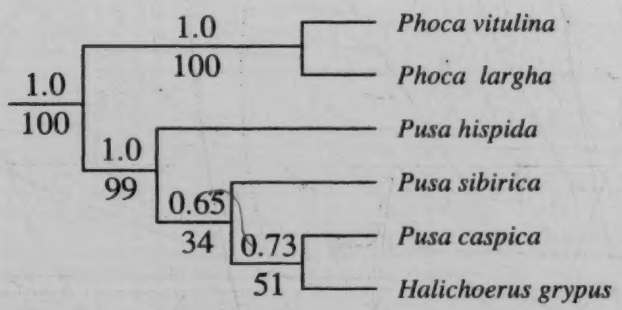


**Figure 5-5 (b).** Phocina topology recovered by combined nuclear (no *GHR*) + mitochondrial (by rate) dataset. Above branch: BPP. Below branch: MLBP.

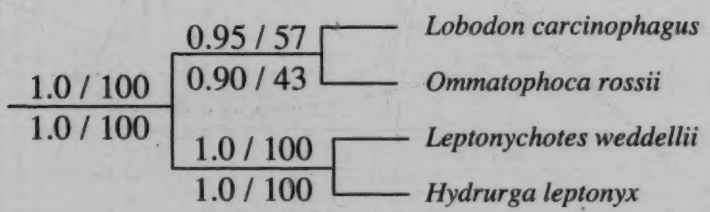


**Figure 5-5 (c).** Lobodontini relationships in combined analyses. Above branch: combined (mt no 3rd positions) BPP/MLBP support. Below branch: combined (mt by rate) BPP/MLBP support.

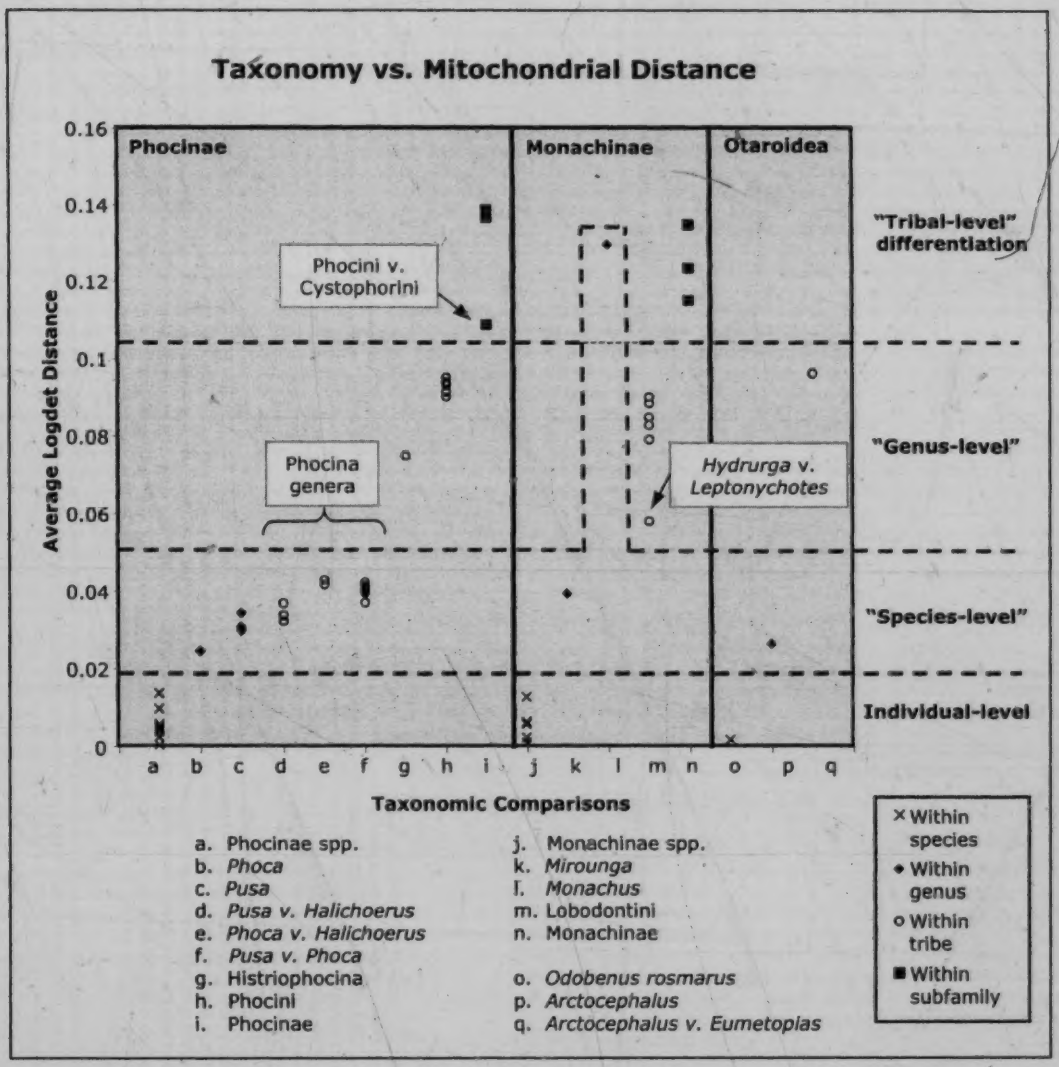


**Figure 5-6.** Taxonomy vs. mitochondrial distance. Taxonomic comparisons were made varying from between individuals within a species (a,j,o), up to between tribes (i,n). Distances were averaged for nested taxonomic comparisons (see Methods). Some points discussed in text are highlighted.

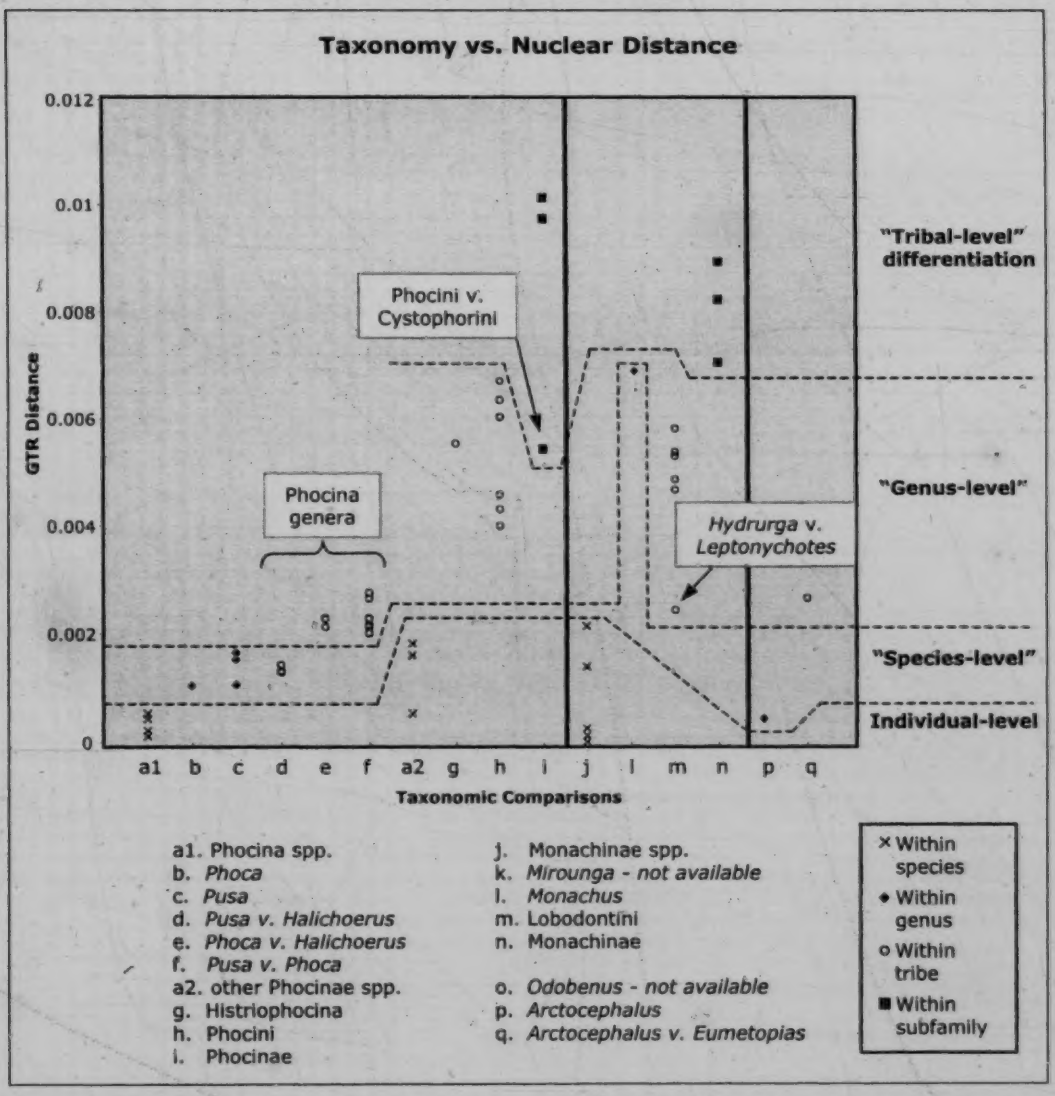


**Figure 5-7.** Taxonomy vs. nuclear genetic distance. Taxonomic comparisons were made varying from between individuals within a species (a1,a2,j,o), up to between tribes (i,n). Distances were averaged for nested taxonomic comparisons (see Methods). Two comparisons were made for mitochondrial distances (Figure 5-6, k,o), but not for nuclear distances. Some points discussed in text are highlighted.

**References**


Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: Selection of bestfit models of protein evolution. Bioinformatics 21, 2104-2105.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel metropolis coupled markov chain monte carlo for bayesian phylogenetic inference. Bioinformatics 20, 407-415.

Amrine-Madsen, H., Koepfli, K.-P., Wayne, R.K., Springer, M.S., 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. Mol.Phylogenet.Evol. 28, 225-240.

Arnason, U., Bodin, K., Gullberg, A., Ledje, C., Mouchaty, S., 1995. A molecular view of pinniped relationships with particular emphasis on the true seals. J.Mol.Evol. 40, 78-85.

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., Lehman, N., Petrov, E.A., Vainola, R., 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol.Phylogenet.Evol. 41, 345-354.

Barker, F.K., Lutzoni, F.M., 2002. The utility of the incongruence length difference test. Syst.Biol. 51, 625-637.

Barriel, V., 1994. Molecular phylogenies and how to code insertion/deletion events. Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie 317, 693-701.

Bengtson, J.L., 2002. Crabeater seal: *Lobodon carcinophagus*. In: Perrin, W.F., Wursig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of marine mammals. Academic Press, San Diego, CA, pp. 302-304.

Berta, A., Sumich, J.L., Kovacs, K.M., 2006. Marine mammals: evolutionary biology. Elsevier Academic Press.

Berta, A., Wyss, A.R., 1994. Pinniped phylogeny. Proc.San Diego Soc.Natl.Hist. 29, 33-56.

Bininda-Emonds, O.R.P., Gittleman, J.L., Purvis, A., 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biol.Rev. 74, 143-175.

Bininda-Emonds, O.R.P., Russell, A.P., 1996. A morphological perspective on the phylogenetic relationships of the extant phocid seals (Mammalia: Carnivora: Phocidae).

Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of Scincid lizards. Syst.Biol. 54, 373-390.

Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst.Biol. 56, 643-655.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst.Biol. 42, 384-397.

Burns, J.J., Fay, F.H., 1970. Comparative morphology of the skull of the Ribbon seal, *HistrioHistriophoca fasciata*, with remarks on systematics of Phocidae. J.Zool.Lond. 161, 363-394.

Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst.Biol. 56, 400-411.

Castoe, T.A., Parkinson, C.L., 2006. Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). Mol.Phylogenet.Evol. 39, 91-110.

Chapskii, K.K., 1955a. An attempt at revision of the systematics and diagnostics of the subfamily Phocinae. Trudy Zool. Inst. Akad. Nauk SSSR 17, 160-199.

Clayton, J.W., Fernando, E.S., Soltis, P.S., Soltis, D.E., 2007. Molecular phylogeny of the tree-of-heaven family (Simaroubaceae) based on chloroplast and nuclear markers. Int. J. Plant Sci. 168, 1325-1339.

Cunningham, C.W., 1997. Can three incongruence tests predict when data should be combined? Mol.Biol.Evol. 14, 733-740.

Darlu, P., Lecointre, G., 2002. When does the incongruence length difference test fail? Mol.Biol.Evol. 19, 432-437.

Davis, C.S., Delisle, I., Stirling, I., Siniff, D.B., Strobeck, C., 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. Mol.Phylogenet.Evol. 33, 363-377.

de Muizon, C., 1982. Phocid phylogeny and dispersal. Ann.S.Afr.Mus. 89, 175-213.

Delisle, I., Strobeck, C., 2002. Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. Mol.Biol.Evol. 119, 357-361.

Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Mol.Phylogenet.Evol. 37, 192-201.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat.Rev.Genet. 6, 361-375.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol.Phylogenet.Evol. 17, 401-406.

Dowton, M., Austin, A.D., 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy - the behavior of the incongruence length difference test in mixed-model analyses. Syst.Biol. 51, 19-31.

Dragoo, J.W., Honeycutt, R.L., 1997. Systematics of Mustelid-like carnivores. J.Mammal 78, 426-443.

Drummond, A., Strimmer, K., 2001. PAL: An object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics 17, 662-663.

Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1995. Constructing a significance test for incongruence. Syst.Biol. 44, 570-572.

Fay, F.H., Rausch, V.R., Feltz, E.T., 1967. Cytogenetic comparison of some pinnipeds (Mammalia: Eutheria). Can.J.Zool. 45, 773-778.

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M.A., 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst.Biol. 54, 317-337.

Flynn, J.J., Nedbal, M.A., 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs incompatibility among multiple data sets. Mol.Phylogenet.Evol. 9, 414-426.

Fulton, T.L., Strobeck, C., 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. Mol.Phylogenet.Evol. 41, 165-181.

Fyler, C.A., Reeder, T.W., Berta, A., Antonelis, G., Aguilar, A., Androukaki, E., 2005. Historical biogeography and phylogeny of monachine seals (Pinnipedia: Phocidae) based on mitochondrial and nuclear DNA data. J.Biogeogr. 32, 1267-1279.

Gibson, A., Gowri-Shankar, V., Higgs, P.G., Rattray, M., 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol.Biol.Evol. 22, 251-264.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst.Biol. 52, 696-704.

Harkonen, T., Harding, K.C., 2001. Spatial structure of harbour seal populations and the implications thereof. Can.J.Zool. 79, 2115-2127.

Hendey, Q.B., 1972. The evolution and dispersal of the Monachinae (Mammalia: Pinnipedia). Ann.S.Afr.Mus. 59, 99-113.

Higdon, J.W., Bininda-Emonds, O.R.P., Beck, R.M.D., H., F.S., 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a mutigene dataset. BMC Evol.Biol. 7, 216.

Hince, B., 2000. The Antarctic dictionary: a complete guide to Antarctic english. CSIRO publishing, Collingwood, Australia.

Ho, J.W.K., Adams, C.E., Bin Lew, J., Matthews, T.J., Ng, C.C., Shahabi-Sirjani, A., Tan, L.H., Zhao, Y., Easteal, S., Wilson, S.R., Jermiin, L.S., 2006. SeqVis: Visualization of compositional heterogeneity in large alignments of nucleotides. Bioinformatics 22, 2162-2163.

Housley, D.J.E., Zalewski, Z.A., Beckett, S.E., Venta, P.J., 2006. Design factors that influence PCR amplification success of cross-species primers among 1147 mammalian primer pairs. BMC Genomics 7, 253.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Illiger, J.C.W., 1811. Prodomus systematics Mammalium et Avium. C. Salfeld, Berlin, Germany.

Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, A.W.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimates. Syst.Biol. 53, 638-643.

Jiang, Z., Priat, C., Galibert, F., 1998. Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. Mamm.Genome 9, 577-587.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J.Am.Stat.Assoc. 90, 773-795.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066.

King, J.E., 1966. Relationships of the Hooded and Elephant seals (genera *Cystophora* and *Mirounga*). J.Zool. 148, 385-398.

King, J.E., 1983. Seals of the world. British Museum (Natural History) and Oxford University Press, London and Oxford.

Kjer, K.M., Honeycutt, R.L., 2007. Site specific rates of mitochondrial genomes and the phylogeny of eutheria. BMC Evol.Biol. 7, 8.

Ledje, C., Arnason, U., 1996. Phylogenetic relationships within caniform carnivores based on analyses of the mitochondrial 12S rRNA gene. J.Mol.Evol. 43, 641-649.

Lemmon, A.R., Moriarty, E.C., 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst.Biol. 53, 265-277.

Lewis, P.O., Holder, M.T., Holsinger, K.E., 2005. Polytomies and Bayesian phylogenetic inference. Syst.Biol. 54, 241-253.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas III, E.J., Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.A., deJong, P.J., Kirkness, E.F., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.-W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.-P., Parker, H.G., Pollinger, J.P., Searle, S.M.J., Sutter, N., Thomas, R., Webber, C., Platform, B.I.G.S., Lander, E.S., 2005. Genome sequence, comparative analysis and haplotypes structure of the domestic dog. Nature 438, 803-819.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol.Biol.Evol. 11, 605-612.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E., O'Brien, S.J., 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat.Genet. 15, 47-56.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst.Biol. 55, 21-30.

Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary analysis. Version 1.12. http://mesquiteproject.org

McLaren, I.A., 1960a. On the origin of the Caspian and Baikal seals and the paleoclimatological implication. Am.J.Sci. 258, 47-65.

Mueller, R.L., Macey, J.R., Jaekel, M., Wake, D.B., Boore, J.L., 2004. Morphological homoplasy, life history evolution, and historical biogeography of Plethodontid salamanders inferred from complete mitochondrial genomes. Proc.Natl.Acad.Sci.U.S.A. 101, 13820-13825.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614-618.

Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society. Series B (Methodological) 56, 3-48.

Nishihara, H., Okada, N., Hasegawa, M., 2007. Rooting the eutherian tree: the power and pitfalls of phylogenomics. Genome Biol. 8, R199.

Nylander, J.A.A., 2004. MrModeltest. Version 2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst.Biol. 53, 47-67.

Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12, 357-358.

Palo, J.U., Makinen, H.S., Helle, E., Stenman, O., Vainola, R., 2001. Microsatellite variation in ringed seals (*Phoca hispida*): genetic structure and history of the Baltic Sea population. Heredity 86, 609-617.

Palo, J.U., Vainola, R., 2006. The enigma of the landlocked Baikal and Caspian seals addressed through phylogeny of phocine mitochondrial sequences. Biol.J.Linn.Soc. 88, 61-72.

Pecon Slattery, J., Murphy, W.J., O'Brien, S.J., 2000. Patterns of diversity among SINE elements isolated from three Y-chromosome genes in Carnivores. Mol.Biol.Evol. 17, 825-829.

Posada, D., Crandall, K.A., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Quicke, D.L.J., Jones, O.R., Epstein, D.R., 2007. Correcting the problem of false incongruence due to noise imbalance in the incongruence length difference (ILD) test. Syst.Biol. 56, 496-503.

Rambaut, A., 2002. Se-Al. v.2.0a11. http://tree.bio.ed.ac.uk/software/seal/

Rambaut, A., Drummond, A.J., 2007. Tracer. v1.4. http://tree.bio.ed.ac.uk/software/tracer/

Rannala, B., 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. Syst.Biol. 51, 754-760.

Repenning, C.A., Ray, C.E., 1977. The origin of the Hawaiian monk seal. Proceedings of the Biological Society of Washington 89, 667-688.

Repenning, C.A., Ray, C.E., Grigorescu, D., 1979. Pinniped biogeography. In: Boucot, A.J. (Ed.), Historical biogeography, plate tectonics, and thee changing environment: proceedings of the thirty-seventh annual biology colloquium and selected papers. Oregon State University Press, pp. 357-369.

Rodriguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst.Biol. 56, 389-399.

Rogers, T.L., 2002. Leopard seal: *Hydrurga leptonyx*. In: Perrin, W.F., Wursig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of marine mammals. Academic Press, San Diego, CA, pp. 692-693.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Sasaki, H., Numachi, K., Grachev, M.A., 2003. The origin and genetic relationships of the Baikal seal, *Pusa sibirica*, by restriction analysis of mitochondrial DNA. Zool.Sci. 20, 1417-1422.

Sato, J.J., Hosoda, T., Wolsan, M., Suzuki, H., 2004. Molecular phylogeny of Arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of the ferret-badgers and skunks. Zool.Sci. 21, 111-118.

Scheffer, V.B., 1958. Seals, sea lions, and walruses: a review of the Pinnipedia. Stanford University Press, Stanford, California.

Shaughnessy, P.D., Fay, F.H., 1977. A review of the taxonomy and nomenclature of North Pacific Harbour seals. J.Zool.Lond. 182.

Stamatakis, A., 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. Proceedings of IPDPS2006, Rhodos, Greece.

Stamatakis, A., 2006b. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Stanhope, M.J., Czelusniak, J., Si, J.-S., Nickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Binding Protein, with convincing evidence for bat monophyly. Mol.Phylogenet.Evol. 1, 148-160.

Sullivan, J., 1996. Combining data with different distribution of among-site rate variation. Syst.Biol. 45, 375-380.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sunderland, Massachusetts.

Teeling, E.C., Scally, M., Kao, D.J., Romagnoll, M.L., Springer, M.S., Stanhope, M.J., 2000. Molecular evidence regarding the origin of echolocation and flight in bats. Nature 403, 188-192.

Venta, P.J., Brouillette, J.A., Yuzbasiyan-Gurkan, V., Brewer, G.J., 1996. Gene-specific universal mammalian sequence-tagged sites: application to the canine genome. Biochem.Genet. 34, 321-341.

Westlake, R.L., O'Corry-Crowe, G.M., 2002. Macrogeographic structure and patterns of genetic diversity in harbor seals (*Phoca vitulina*) from Alaska to Japan. J.Mammal. 83, 1111-1126.

Wheeler, W.C., Pickett, K.M., 2008. Topology-Bayes versus clade-Bayes in phylogenetic analysis. Mol.Biol.Evol. 25, 447-453.

Wyss, A.R., 1987. The walrus auditory region and the monophyly of the pinnipeds. Am.Mus.Novit. 2871, 1-31.

Wyss, A.R., 1988. On "retrogression" in the evolution of the Phocinae and phylogenetic affinities of the monk seals. Am.Mus.Novit. 2924, 1-38.

Wyss, A.R., Flynn, J.J., 1993. A phylogenetic analysis and definition of the Carnivora. In: Szalay, F., Novacek, M., McKenna, M. (Eds.), Mammal Phylogeny: Placentals. Springer-Verlag, New York, pp. 32-52.

Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Syst.Biol. 50, 408-424.

## Chapter 6

## Multiple fossil calibrations and nuclear loci provide new insight into true seal biogeography and divergence times

### Introduction

The historical biogeography of the true seal family (Phocidae, Pinnipedia) has been debated throughout the last century, but hypotheses vary widely and little consensus has been reached. The debate concerning the historical classification of pinnipeds as monophyletic (Illiger, 1811; Flower and Lyddekker, 1891; Gregory and Hellman, 1939; Simpson, 1945) versus the resurrected theory of diphyly (Mivart, 1885) during the 1960s to 1980s, impacted the interpretation of the phylogenetic affinity of early pinnipeds, Enaliartidae, and the initial origins of Pinnipedia. Based on a diphyletic origin, an Atlantic (or Palearctic) origin was proposed for the Phocidae and a Pacific origin for the Otariidae (fur seals and sea lions) and Odobenidae (walrus), often including the Enaliarctidae (McLaren, 1960b; Mitchell and Tedford, 1973; Ray, 1976; Repenning et al., 1979). Pinnipedia is now known to be a monophyletic group that is sister to the Musteloidea within the arctoid Carnivora (Flynn et al., 2005; Fulton and Strobeck, 2006; Sato et al., 2006) and the Enaliarctidae are considered to be the earliest pinnipeds, not otarioids (Berta et al., 1989, see Figure 6-1). Most recent hypotheses support a Pacific origin of pinnipeds, continued early otariid and odobenid evolution in the Pacific and movement of the phocid ancestor through the Central American Seaway, resulting in an Atlantic origin for Phocidae (Bininda-Emonds and Russell, 1996; Demere et al., 2003; Fyler et al., 2005). An alternate hypothesis of a southern North American non-marine pinniped origin has also recently been proposed (Arnason et al., 2006), with otarioids dispersing west (including enaliarctids as otarioids) into the Pacific and phocids eastward into the Atlantic.

With pinniped monophyly reaffirmed, the extinct north Pacific Desmatophocidae have been identified as belonging to the Phocidae (Wyss and Flynn, 1993; Berta and Wyss, 1994). However, the proposal of an Odobenidae-Phocidae grouping (Wyss, 1987, 1988; Wyss and Flynn, 1993; Berta and Wyss, 1994), as opposed to the historical and molecular-supported Odobenidae-Otariidae grouping Otarioidea (Repenning et al., 1979; King, 1983; Ledje and Arnason, 1996; Dragoo and Honeycutt, 1997; Flynn and Nedbal, 1998; Davis et al., 2004; Delisle and Strobeck, 2005; Flynn et al., 2005; Arnason et al., 2006; Fulton and Strobeck, 2006; Higdon et al., 2007), leads to some difficulty in interpreting the placement of the earliest otarioids, the

Imagotariinae. First described as a subfamily of Otariidae (=Otarioidea) (Mitchell, 1968), they have been recognized as paraphyletic and associated with the Odobenidae (Demere, 1994; Kohno, 1994, see Figure 6-1). Otariids were not included in the 1994 studies, but recent work including both Otariidae and Phocidae groups the "Imagotariinae" within Odobenidae to the exclusion of Otariidae, leaving basal Otariidae-Odobenidae-Phocidae relationships unresolved (Kohno, 2006).

The family Phocidae is comprised of two subfamilies, Monachinae ("southern seals") and Phocinae (northern seals), and each is divided into three tribes (Figure 6-1). Prior to King (1966), a third subfamily, Cystophorinae (bladder-nosed seals), was used, including the northern seal, *Cystophora cristata* and the elephant seals, *Mirounga*; further complicating early biogeography and fossil description. Though the two-subfamily system is near-universally accepted, the only recent comprehensive analysis of many phocid fossil taxa was framed within the three subfamily system (Koretsky, 2001). The present tribal system as proposed by Burns and Fay (1970) is strongly supported by recent molecular work (Davis et al., 2004; Arnason et al., 2006; Higdon et al., 2007, Chapter 5). So, while extensive lists of existing fossils have been compiled (Miyazaki et al., 1994; Demere et al., 2003), rigorous phylogenetic study including fossil taxa remains to be performed within the present, well-defined taxonomic system.

Several recent studies have employed molecular dating techniques to address aspects of phocid biogeography and divergence times using mitochondrial (mt) RFLPs (Sasaki et al., 2003), mtDNA (Arnason et al., 2006; Palo and Vainola, 2006), or combined nuclear and mtDNA (Fyler et al., 2005; Higdon et al., 2007). All but one of these studies (Sasaki et al., 2003), have employed some method of relaxing the molecular clock with one or more fossil calibration points. Because of the different focus of each study, the calibration times used have been variable, though all of these studies are united by their use of primarily or exclusively mtDNA.

Just as a range of methods can be used to relax the molecular clock (Welch and Bromham, 2005), various methods of enforcing fossil calibrations can be employed. Fossils can be used to calibrate the clock across the tree by fixed-point calibrations or through a number of different techniques of creating hard or soft bounds on divergence times (Yang and Rannala, 2006). Point calibrations are difficult to properly implement if uncertainty exists regarding the age of the fossil or its placement on the tree. Assignment of hard minimum bounds allows fossils to represent minimum divergence times and such bounds are often implemented with a very liberal upper bound or with no upper bound at all. While this allows flexibility in assigning a molecular date to the node in question (as opposed to a point estimate), it has been shown that

using only minimum limits can lead to overestimated divergence times (Hugall et al., 2007). Recent Bayesian techniques allow fossil calibrations to be implemented as prior distributions that can more accurately reflect the uncertainty involved (Drummond et al., 2006; Yang and Rannala, 2006).

In this study, I use 16 nuclear markers in combination with mtDNA to examine the effect of including nuclear DNA and the use of multiple markers on the estimated divergence times. Multiple soft-bound fossil-based calibrations are implemented as normal, lognormal, gamma, or exponential prior distributions (see Drummond et al., 2006; Ho, 2007) in a relaxed clock framework using the program BEAST (Drummond and Rambaut, 2007). The use of multiple DNA markers spanning a variety of evolutionary rates, combined with multiple calibration points provides a framework for evaluating previous biogeographic hypotheses and examining the potential associations of some fossils of uncertain phylogenetic placement within the presently estimated topological framework and divergence times.

**Methods**

16 nuclear gene fragments of varying evolutionary rate were selected for phylogeny reconstruction and evolutionary divergence time estimation. Sequences were obtained as described in Chapter 5, and the loci used are listed in Table 5-4 (but not including *GHR*). In addition, 12 mt protein coding genes and 12S rRNA are used, with the 3$^{rd}$ codon position bases excluded due to their significant base composition bias across taxa (see Chapter 5). A likelihood ratio test for clock-like evolution (Felsenstein, 1981) was performed using PAUP* v.4.0b10 (Swofford, 2003) for both the combined nuclear and the mtDNA datasets. For both tests including all taxa and those tests including only the ingroup taxa (Phocidae), a molecular clock was strongly rejected (p=0.000).

Bayesian estimation of divergence times was performed using BEAST v.1.4.7 (Drummond and Rambaut, 2007) for this nuclear + mtDNA (3$^{rd}$ positions excluded) dataset. Each gene was allowed its own independent evolutionary model and parameters by manually editing the XML control file produced by BEAUti v.1.4.7. Models were selected using AIC selection in MrModelTest (Nylander, 2004), a restricted version of ModelTest (Posada and Crandall, 1998). HKY (Hasegawa et al., 1985) was selected for *RAG2*; HKY+Γ for *ADORA3*, *APOB*, *PNOC*, *FLVCR1*, and *RHO*; HKY+I for *BDNF* and *CREM*; HKY+I+Γ for *IRBP*; GTR+Γ for *BRCA1* and *PLCB4*; GTR+I for *FES* and *APP*; and GTR+I+Γ for mtDNA. Base frequencies

were set to be estimated for all partitions, except for *CHRNA1* and *RAG1*, where the K80+Γ (Kimura, 1980) and SYM+I+Γ (Zharkikh, 1994) models were implemented, respectively. A Yule process tree prior was used and rate variation across branches was uncorrelated and lognormally distributed. Tuning parameters for the MCMC operators were set to auto-optimize and successive runs were tuned accordingly. Each MCMC chain was started from a random tree and run for 30 000 000 generations. Three independent runs were performed; each run was sampled every 1000 generations and 10% of samples were removed from each run as burn-in. The runs were combined using LogCombiner v.1.4.7 to obtain a number of independent samples from the marginal posterior distribution (ESS, Effective Sample Size) greater than 200, determined using Tracer v1.4 (Rambaut and Drummond, 2007). The analysis was run without data to determine that priors were being implemented properly (and not interacting unexpectedly) and that the data was informative (i.e. the posterior values with data are different than those without data) to ensure that the final results were not solely the result of the priors (Drummond et al., 2006).

Fossil calibrations were implemented as normal, lognormal, or gamma-distributed priors and are listed in Table 6-1. A tree prior was used for all remaining nodes. All fossil-calibrated priors represent soft-bounded priors to allow for possible uncertainty in fossil dates or node assignment. Reasoning for each calibration is as follows, with calibration numbers referring to Table 6-1.

1. Feliform-caniform divergence (Carnivora most recent common ancestor, MRCA). Based on the work of Wesley-Hunt and Flynn (2005), the earliest caniforms appeared ~45-40 million years ago (Ma) setting a lower bound, and the earliest viverravids, early carnivoramorphs representing a stem lineage before the crown-clade Carnivora diverge (Flynn and Wesley-Hunt, 2005), appeared ~65-60 Ma, setting an upper bound.
2. Canidae-Arctoidea divergence (Caniformia MRCA). The earliest known canid, *Hesperocyon gregarius*, appeared in the late Eocene of North America, possibly as early as 39.5 Ma (Wang, 1994; Munthe, 1998). The more primitive *Prohesperocyon* has been found between 37.1 and 35.5 Ma (Munthe, 1998). *Mustelavus* is an early arctoid (Wang et al., 2005) and was contemporaneous with *Hesperocyon* in the late Eocene (~37-33 Ma) (Wang, 1994). Based on these, a mean divergence was set just before the appearance of the two lineages (40 Ma), with soft upper and lower bounds corresponding roughly with the more recent possible assigned age for the earliest

fossils (~35 Ma) and an upper bound (~45 Ma) with the appearance of caniforms (see calibration 1).

3. Musteloidea-Pinnipedia divergence (Mustelida MRCA). The earliest pinniped, *Enaliarctos tedfordi*, appeared ~28-25 Ma in the Yaquina formation in Oregon (Berta, 1991). Early mustelids are difficult to place phylogenetically, but *Promartes* (Oligobuninae) is generally placed within Mustelidae (Wang et al., 2005; Finarelli, 2008) and may be found as early as ~27 Ma (Baskin, 1998). The appearance of both of these genera represents a minimum divergence time for mustelids and pinnipeds.
4. Otariidae-Odobenidae (Otarioidea MRCA). The ~16 Ma fossil taxon *Proneotherium repenningi* (Imagotariinae) likely represents a basal odobenid (Kohno et al., 1994; Demere et al., 2003) and was used as a broad calibration for the divergence between these families. The sister group to the Odobeninae (including the extant walrus, *Odobenus rosmarus*) is the Dusignathinae, but species belonging to the Dusignathinae are not observed until ~ 6.7-5.2 Ma (Demere et al., 2003).

   As there is some uncertainty regarding the affinity of the Imagotariinae, a second analysis was performed with this constraint based on early imagotariines removed.
5. Monachinae-Phocinae (Phocidae MRCA). Both the monachine *Monotherium? wymani* and the phocines *Leptophoca lenis* and *Prophoca* sp. indet. were found in the western north Atlantic ~16.3 -14.2 Ma (Repenning et al., 1979; Demere et al., 2003). A prior distribution with 95% range of 28.28-14.7 Ma and median age 17.6 Ma was used to calibrate this node.
6. *Halichoerus-Phoca-Pusa* (Phocina MRCA). Several extant Phocina species are observed by the late Pleistocene, 0.79-0.01 Ma, and a *Phoca* species c.f. *P. vitulina* has been found, dated to the early Pleistocene, 1.64-0.79 Ma (Demere et al., 2003). The divergence of these genera was set to a median age of ~5 Ma, but with a large 95% interval to encompass uncertainty, from ~23.7-1.4 Ma.
7. Lobodontini-Miroungini MRCA. Although they have not been subjected to recent rigorous phylogenetic study, species of *Acrophoca* have been associated with the Lobodontini, specifically *Hydrurga leptonyx* (de Muizon, 1982). *Acrophoca* species found in Peru are most likely ~7-5 Ma (de Muizon and DeVries, 1985). *Homiphoca capensis* has also been allied with the lobodontines (de Muizon, 1982) and is of the same age (5.2-3.4 Ma) as *Callophoca obscura*, a member of the Miroungini (de Muizon, 1982; Demere et al., 2003). Because of the uncertain placement of

*Acrophoca* within Lobodontini (i.e. stem or crown lineage), the 95% interval was set to 15.75-6.8, with a median age of ~10 Ma for the divergence between these tribes.

8. Lobodontini MRCA. As mentioned for calibration 7, *Homiphoca capensis* and *Acrophoca longistirostris* are associated with different extant lobodontine genera, *Lobodon* and *Hydrurga*, respectively. *H. capensis* has been hypothesized to be a potential ancestor of *Lobodon* (Muizon and Hendey, 1980; de Muizon, 1982). Although *H. capensis* is not found until the early Pliocene, it has been described from South Africa (Hendey and Repenning, 1972) and mentioned to have been found in the eastern USA (Ray, 1976) of the same period. Therefore, the divergence among extant lobodontine genera was set to have a soft minimum for the oldest possible known age for this species.

Analyses were also run for the nuclear and mtDNA separately. For these, two runs of 15 000 000 generations were combined to achieve ESS values >200 for all parameters. Samples were taken every 1000 generations and 10% of these were removed for burn-in.

To ensure that the root age (Caniformia-Feliformia split) and deep node calibrations (canid-arctoid and musteloid-pinniped divergences) were not heavily influencing the other divergence times, a second set of priors was used. A normally distributed prior was applied to the caniform-feliform divergence (root node) with a mean age of 65.0 Ma (the earliest possible appearance of viverravids) and standard deviation of 5.0 [5% 56.78, 95% 73.22]. The Caniformia and Arctoidea calibrations, to which fossil-based priors were previously applied, were set to have a tree prior. Two runs of 30 000 000 were performed and combined as above.

## Results and Discussion

### *Nuclear vs. mitochondrial divergence time estimates and use of multiple fossil calibrations*

Analyses of nuclear and mt DNA separately generally recovered similar divergence times (Table 6-2), though some estimates were 2-3 times different from one another without overlap in the 95% highest posterior density interval (95% HPD). In general, family-level nodes and higher were estimated to be older using nuclear DNA, while more recent nodes (within families) were estimated to be older using mtDNA. This is unsurprising, as mtDNA mutates more quickly than most of the nuclear genes employed, accumulating more mutations near the tips of the tree, but

leading to saturation at the deeper nodes and an underestimation of these divergence times (e.g. see Hugall et al., 2007). When the two datasets were combined (mtDNA excluding 3rd codon positions + nuclear DNA), the recovered dates were quite similar to those estimated by nuclear DNA alone, but the more recent clades (within subfamilies) had slightly older divergence times (Tables 6-2, 6-3). To include both the information about recent divergences (i.e. species divergences) provided by the mtDNA and the information for deeper nodes (i.e. tribes to superfamilies) provided by the more slowly evolving nuclear DNA, the combined dataset is used to infer biogeography and the recovered topology is shown in Figure 6-2. The nuclear-only and mt-only analyses recovered different topologies than the combined analysis for the Phocina and Lobodontini as discussed in Chapter 5.

While the difference between mt and nuclear estimations is not unexpected, it has important implications for comparisons with previous studies that primarily utilized mtDNA. The dates recovered from the nuclear + mt analysis (Table 6-3) for the divergences within the two Phocidae subfamilies are most similar, though slightly more recent, than the other study that included both nuclear and mtDNA using a supertree dating technique based on several fossil calibrations and both nuclear and mt genes (Higdon et al., 2007), though their supermatrix estimated dates were generally about twice as old as those estimated here. Similarly, dates within subfamilies recovered here also tend to be younger than those obtained by other studies that used primarily mtDNA and only calibrated the root node (Fyler et al., 2005; Arnason et al., 2006; Palo and Vainola, 2006). As the dates recovered in this study correspond closely to those used as calibrations in other studies (Arnason et al. 2006: feliform-caniform 52 Ma; Fyler et al. 2005: Phocinae-Monachinae 15 and 17 Ma; Palo and Vainola 2006: feliform-caniform ~55 Ma), the differences between this study and those previous may be attributed to the DNA marker type used (nuclear or mt), the relaxed clock method employed and the program in which it was implemented, or to the number of fossil calibrations enforced. As these studies implemented several different relaxed clock methods, I did not examine the effect of the program and method employed, but instead examined the effect of using mt versus nuclear data within the same relaxed clock framework. As discussed above, the mtDNA estimated dates for the more recent nodes were considerably older than those estimated using nuclear DNA (Table 6-2). My estimated mtDNA divergence times within the Monachinae are comparable to those obtained by Fyler et al. (2005), who used 1 nuclear and 3 mt genes. The mt-only dates are slightly younger across both Monachinae and Phocinae than those obtained by Arnason et al. (2006) in their analysis of complete mt amino acid sequence. This difference may be attributable to the use of

amino acid sequence versus DNA sequence, the exclusion of mt 3rd position bases in this study, or to the use of different relaxed clock methods. However, given that all analyses utilizing primarily mtDNA are more comparable to the mtDNA-only analysis than to the nuclear-only analysis performed here, it is reasonable that most of the discrepancies between the present and previous studies are due to the different evolutionary patterns of mtDNA compared to a combination of several independent nuclear loci. As 16 nuclear loci plus mtDNA were utilized in this study, the combination of many genes of differing evolutionary rate may be more reflective of the rate of evolution of the whole genome, as opposed to analysis of mtDNA alone. Including multiple loci may yield a better estimation of the species tree than the mt gene tree analyzed alone.

Another large contributor to the difference between this and previous studies is the number of fossil calibrations used. The dates estimated by this study are particularly different from those obtained by Arnason et al. (2006). Though I obtain a very similar divergence time as was used to calibrate their tree, in this study, calibrations were placed across the tree to allow calibration at multiple levels, locally "anchoring" the clock (Benton and Donoghue, 2007). When an older prior was enforced on the root node to test the effect of the root age calibration, the deeper nodes were estimated to be unreasonably old: Carnivora ~76 Ma, Caniformia ~75 Ma, and Arctoidea ~58 Ma. However, estimated dates within Pinnipedia were virtually unchanged. The estimated divergence between Otarioidea and Phocidae was slightly older (~24.9 Ma) than the estimate using more reasonable root ages (Table 6-3). Estimated divergence times within Phocidae were very similar and those within Monachinae were slightly older (<1 million year difference) and well within the 95% HPD of the 'reasonably' aged root node. This illustrates the benefit of using multiple calibrations across the tree. All the nodes prior to the next calibration (at Pinnipedia) being pulled deeper back in time with the older root node calibration (and removal of the Caniformia and Arctoidea calibrations) is consistent with other findings that the use of a single deep external calibration can yield inflated divergence times (Hugall et al., 2007). My results using multiple calibrations are most similar to the supertree of Higdon et al. (2007), who also used multiple calibrations and incorporated some nuclear information. The large difference with their supermatrix dates must be methodological, but is discussed in their manuscript (Higdon et al., 2007). In studies enforcing only a single root calibration, the present results are more similar to those enforcing a calibration very close to the group of interest (Fyler et al., 2005), than to that which enforces a much deeper single calibration (Arnason et al., 2006). Use of a single calibration in most previous studies or using only minimum constraints (i.e. Higdon et al. 2007),

combined with the primary use of mitochondrial DNA, likely led to inflated ingroup divergence times compared to those estimated here.

*Origin of Pinnipedia, Otarioidea, and Phocidae*

The root node (feliform-caniform split) was estimated using a lognormally distributed prior with a 95% range from 40.865 to 63.22 Ma. The recovered age was ~54 Ma. The divergence times for both the Arctoidea and Caniformia were estimated to be at the oldest end of the prior distributions for both nodes (Table 6-1) and were 39.412 Ma and 50.313 Ma, respectively (Table 6-3). The divergence between pinniped lineages Phocidae and Otarioidea was not constrained and was estimated to be ~22.1 Ma. Members of the early pinniped family Enaliarctidae have been found in Oregon in the northeastern Pacific Ocean, dated ~28-25 Ma (Berta, 1991). This family persisted across the North Pacific until at least ~16 Ma (Miyazaki et al., 1994; Demere et al., 2003). The "Oligocene seal" of indeterminate genus and species, discovered in South Carolina dated ~27-24 Ma, was assigned to the Phocidae (Koretsky and Sanders, 2002). However, only two partial femora were described and few other carnivore taxa were used for comparative purposes (Koretsky and Sanders, 2002). It is inconsistent with the finding of this study that the Oligocene seal belongs to the Phocidae, as its occurrence predates both the oldest known otarioids, the Imagotariinae, and the oldest known phocids, the Desmatophocidae, neither of which is found until the early Miocene (Demere et al., 2003). It is also possible that the fossil was correctly assigned to the Phocidae, but its estimated age may be incorrect (Demere et al., 2003). The ~22 Ma divergence between Phocidae and Otarioidea is similar to Higdon et al.'s supertree estimate of ~23 Ma, but very different than the ~33 Ma age estimated by Arnason et al. (2006). On the basis of Arnason et al.'s estimated date, they postulate a non-marine pinniped origin in southern North America, including the Oligocene seal in the Phocidae and the Enaliarctidae in Otarioidea. While my estimates provide no support for or against a non-marine origin, placement of the Enaliarctidae within Otarioidea (originally based on pinniped diphyly) is not supported. However, as I estimate nearly 10 million years (MY) between the divergence of Pinnipedia from Musteloidea and the first-known occurrence of pinnipeds, an origin either in the north Pacific, possibly near present-day Oregon (where the first enaliarctids are found) or more southernly (if the Oligocene seal represents an early pinniped, not phocid) is plausible.

The oldest known record of the phocid sister group, Desmatophocidae, is that of *Desmatophoca brachycephala* from the Astoria formation in Washington, USA ~23-20 Ma (Barnes, 1987). The occurrence of *D. brachycephala* coincides approximately with the molecular estimate for the Phocidae-Otarioidea divergence of 22.1 Ma and is well within the 95% HPD interval (19.769-24.739). The molecular estimate is consistent with the Desmatophocidae belonging to the Phocoidea with the Phocidae (Wyss and Flynn, 1993; Berta and Wyss, 1994). While the desmatophocids were evolving and *Allodesmus* species began to disperse across the northern Pacific (Miyazaki et al., 1994), the unknown phocid ancestor likely traveled through the Central American Seaway between North and South America to reach the eastern USA coast where phocid fossils are first found (Bininda-Emonds and Russell, 1996; Demere et al., 2003; Fyler et al., 2005).

The divergence between the two otarioid families, Otariidae and Odobenidae, was calibrated using a prior distribution ranging from 14.9 to 23.7 Ma (Table 6-1) and estimated to be ~15.1 Ma (14.2-16.3). When no calibration is applied to this node, the estimated date is much more recent (~10.7 Ma, 7.6-13.5). Other nodes were virtually unaffected by the removal of this calibration, except the otarioid-phocid split, which was estimated to be ~1.5 MY more recent. The oldest known members of the Imagotariinae, *Proneotherium repenningi* and *Prototaria*, existed ~16-15 Ma in the northeast Pacific (Kohno et al., 1994; Miyazaki et al., 1994; Kohno, 2006). The Imagotariinae are generally considered to be a paraphyletic assemblage within the Odobenidae (Demere, 1994; Kohno, 1994; Kohno et al., 1994; Demere and Berta, 2001; Kohno, 2006), but were first described as a subfamily within Otariidae (= Otarioidea), as were the extant Otariinae and Odobeninae (Mitchell, 1968; Mitchell and Tedford, 1973). In the present (molecular-only) context, it is difficult to determine whether the Imagotariinae represent early odobenids or otarioids. Early fossils are found within the range of molecular dates estimated in this study, but the prior constraint placed on these molecular dates are somewhat based on these early fossils so any inference about their phylogenetic affinity becomes circular. However, as the molecular dates tend to push the lower bounds of the calibration and are younger yet when uncalibrated, it is possible that at least some of the Imagotariinae (*Prototaria, Proneotherium, Neotherium*) may be stem otarioids as opposed to odobenids. In his initial description of the Imagotariinae, Mitchell (1968) raiseed the possibility that the otarioid ancestor may have been more "walrus-like" than "sea lion-like" and that otariids arose from a stock "paralleling" the Imagotariinae. However, the phocoids *Desmatophoca* and *Allodesmus* were considered their own otarioid subfamilies at that time, complicating interpretation. In contrast, the Dusignathinae are

generally unquestioned as the sister group to the extant Odobenidae (Odobeninae) (Demere, 1994; Kohno, 1994; Demere et al., 2003; Kohno, 2006). The earliest known Dusignathinae are only ~5-8 million years old (Demere et al., 2003; Kohno, 2006), and the early otariid, *Pithanotaria starrii* occurs ~12-7 Ma (Miyazaki et al., 1994; Demere et al., 2003) which are consistent even with the unconstrained estimated divergence time between otariids and odobenids of ~13.5-7.6 Ma (mean 10.7). A better understanding of the relationships and timing within Otarioidea will require including many more species and more calibrations within the superfamily, to better determine the placement of the Imagotariinae and relationships of extant taxa.

*Phocidae*

Fossils assigned to the subfamilies Phocinae and Monachinae are first found together in the western north Atlantic ~15 Ma (Repenning et al., 1979), at the time of the Mid-Miocene thermal maxima (Lyle et al., 2008). These fossils were used to calibrate the divergence of these subfamilies, implemented as lognormal distribution of median age 16.9 Ma and a soft lower bound (the 95% range limit on the prior) of 14.9 Ma. A small 95% HPD interval was recovered, ranging from 14.3-15.2 Ma (mean 14.7), at the low end of the prior distribution of the calibration. This recovered age corresponds to the 15 Ma calibration employed by Fyler et al. (2005), making their divergence estimates comparable (they also used a 17 Ma calibration). This molecular divergence estimate is of the same age as the first fossils, despite more prior weight having been placed on an earlier molecular divergence. If the molecular estimate is accurate, this would imply that the two subfamilies split ~15 Ma in the vicinity of what is now Maryland and Virginia, USA and the known fossils represent the earliest members of each subfamily (Figure 6-3). *Kawas benegasorum* of Argentina is slightly younger than the Calvert formation phocines (dated ~14-12 Ma), and has been assigned as one of the basal members of the Phocinae (Cozzuol, 2001). This species may represent an early lineage that dispersed south and became extinct, given the present lack of other phocine fossils found as far south. An early phocid, *Desmatophoca claytoni*, has been described from the central Paratethys ~15 Ma (Koretsky and Holec, 2002) and placed in its own subfamily basal to both extant subfamilies, though interpretation was under the assumption of three extant subfamilies (including Cystophorinae). However, a molecular divergence between the two extant subfamilies roughly synchronous with the occurrence of this fossil is consistent with a basal placement within Phocidae (Koretsky and Holec, 2002). *D. claytoni* could represent

the first of a series of phocid dispersals across the Atlantic. After the split between Monachinae and Phocinae, the late middle Miocene (~14.6-11 Ma) represents a time of high dispersal for phocids, as fossils have been found across the north Atlantic, Mediterranean, Paratethys, and as far south as Argentina.

*Phocinae: tribal origins and association with ice*

After both the Monachinae and Phocinae have spread across the Atlantic, the first phocine tribe (Erignathini) is estimated to have emerged at ~11.3 Ma (Table 6-3, node 6). This corresponds to a possibly stable exchange between the Arctic and Atlantic oceans beginning <11.5 Ma (Haley et al., 2008). After a long temporal gap, the Cystophorini diverge ~5.4 Ma (4.0-6.7), coincident with the proposed opening of the Bering Straight ~5.5-5.4 Ma (Gladenkov et al., 2002). Both of these Arctic oceanic events would have caused a shift in oceanic conditions, and the coincidental timing (Figure 6-2) suggests that these may have been factors in Erignathini and Cystophorini evolution. Mineral analyses suggest that perennial ice cover may have existed in the Arctic starting 13-14 Ma (Darby, 2008; Krylov et al., 2008), before the divergence of the Erignathini. Despite being the oldest extant lineage, the earliest known bearded seal fossils are only, at most, ~2 MY old from England (Harington, 2008). There are no extinct taxa strictly associated with the Erignathini. Perhaps early entry into the Arctic sea ice could explain the absence of fossils associated with this >10MY old lineage.

All phocine species except the harbour seal exhibit some level of pagophilic (ice-loving) behaviour (Berta et al., 2006). However, present and historical behaviour do not necessarily correspond. Even when compared with the Antarctic seals (Lobodontini), no clear pattern, phylogenetic or otherwise, emerges to explain life history traits such as lactation strategy or mating system type (polygyny, etc.), except that polar bear predation has had a significant impact in more recent times (~1 MY) on behaviour in the Phocinae (Stirling, 1977, 1983; Lydersen and Kovacs, 1999; Davis et al., 2008). The white lanugo (fur at birth) of the Phocini has been associated with ice-breeding, and used to indicate that the tribe's ancestor inhabited ice (i.e. McLaren, 1960a; Arnason et al., 2006). Association with ice is not likely to be the only factor in the evolution of this white coat, as the Antarctic Lobodontini retain the dark lanugo shared by their temperate-water monachine relatives. Hooded and bearded seals shed a light grey lanugo *in utero* (Kovacs et al., 1996), but hooded seals are born as "bluebacks", shiny blue-grey on their back and white underneath, while bearded seals are grey to greyish-brown, sometimes with white

patches. It has been suggested that bearded seals are in the process of evolving cryptic natal colouring via natural selection through polar bear predation (Stirling, 1977). Thus, it is possible that the potential for a white lanugo evolved early in phocine evolution and some factor fixed this condition in the Phocini ancestor. The Phocini also share a reduced chromosome number (2N=32) compared to Erignathini and Cystophorini (2N=34)(Arnason, 1974). As noted by Arnason (2006), this change likely arose in a small population, given the slow rate of karyotypic change in marine mammals and the short time in which the Phocini were isolated from the Cystophorini before speciating, here estimated to be only about 1 MY. In this small ancestral population, genetic drift may have been a factor in fixing lanugo colour, possibly compounded by predation by ancient bears or canids. Whatever the cause of the white Phocini lanugo, it has been retained in all species except the non-pagophilic harbour seal (where it is shed *in utero*), including the land-locked Caspian and Baikal seals. If the Phocini ancestor was not associated with ice by the time this character evolved, terrestrial predation would more than likely have hampered the success of the now-prolific Arctic tribe. So although perennial Arctic sea ice was available for all phocines, their time of entry into the Arctic remains speculative, as no phylogenetic correlation with life history exists and the extant phocine seal record is limited to the Pleistocene.

Determination of the relationship of the extant taxa with the fossil phocines may help in determining both the center of origin for phocine diversification and the relative timing of phocines entry into the Arctic. The two earliest known phocine genera (from the eastern USA 16-14 Ma), *Leptophoca* and *Prophoca,* are found in Belgium ~14.6-11 Ma. Between 14.6 and 6.7 Ma, phocines are found in the Paratethys (Demere et al., 2003). Between 5.2 and 3.4 Ma, the same Phocinae genera can be found on both sides of the Atlantic, in Belgium and the east-central USA coast. Many of the Paratethyan species have been specifically associated with the Phocini (Repenning et al., 1979; de Muizon, 1982; Koretsky, 2001), but this is inconsistent with a Phocini-Cystophorini divergence ~5.4 Ma. The one very extensive morphological phylogenetic study (Koretsky, 2001) could not assign many taxa to a tribe and several aspects of the recovered phylogeny are inconsistent with presently accepted relationships between extant taxa. Determining how the temperate-water species, such as the western Atlantic taxa and the Paratethyan taxa, are related to the extant phocines should aid in determining how many times phocine seals acquired pagophilic tendencies.

*Phocini*

The Phocini diverged from the Cystophorini ~5.4 Ma and the subtribes Histriophocina and Phocina split shortly (~1 MY) afterward. As discussed above, this ancestor may have already adapted to life on ice and changing Arctic conditions likely influenced subsequent speciation. After the subtribes diverged from one another, the two Histriophocina species, the ribbon seal (*Histriophoca*) and harp seal (*Pagophilus*), split first, ~3.3 Ma. This is roughly coincident with the closure of the Central American Seaway ~3 Ma. However, this may be unrelated, as the rise of the Isthmus of Panama (closing the Central American Seaway) may only be a minor factor in Northern Hemisphere Glaciation and not a cause (Lunt et al., 2008; Molnar, 2008). Glacioeustatic-forced allopatric speciation during the Pleistocene has been hypothesized for the split between the Histriophocina species (Davies, 1958; Demere et al., 2003). I estimate that this speciation event occurred earlier (in the late Pliocene), before major glaciation had begun. However, given their present distributions (Figure 6-4), the ribbon seal likely evolved in the North Pacific and the harp seal in the North Atlantic, with this separation reinforced by glaciation. By the Late Pleistocene, the first fossils of each are found in these respective locations (Demere et al., 2003).

The phylogenetic relationships between Phocina species have been debatable and remain to be confidently resolved (Chapter 5). Similar to the maximum likelihood analysis in Chapter 5, the Bayesian inference method employed here recovers the grey seal, *Halichoerus grypus*, as the first branching lineage ~1.9 Ma, followed by the divergence between *Phoca* and *Pusa* ~1.5 Ma (Figure 6-2, Table 6-3). This is the most recent estimate for the divergences within this clade, with others ranging from ~30-60% older (Palo and Vainola, 2006; Higdon et al., 2007) to over twice as old (Arnason et al., 2006). Though the topologies differ, the estimated times for the basal split within Phocina are largely overlapping between this study and that of the Higdon et al. (2007) supertree, indicating a rougly 2 Ma divergence for Phocina.

Arctic water was rapidly cooled with the Northern Hemisphere Glaciation ~2.5 Ma (Lyle et al., 2008). Isotope analyses indicate a regional climatic shift in the Canadian Basin 1-2 Ma and that central Arctic glacial/interglacial cyclicity and pronounced Arctic climate change also occurred ~1 Ma (Haley et al., 2008). Within this same short time frame, nearly all of the Phocina species rapidly radiate (Table 6-3), indicating these Arctic shifts were influential in Phocina evolution. Between 1.64 and 0.79 Ma, fossils resembling *Phoca vitulina* are found in Oregon, USA, and by the Late Pleistocene (0.79-0.01 Ma), *Phoca* species are found in both the north

Pacific and Atlantic, including the Champlain Sea (present-day Ontario and Quebec, Canada). All other fossils of extant species are not found until the late Pleistocene, generally close to the locations they occur in today (Demere et al., 2003).

A Greenland Sea / Barents Sea centre for Phocina divergence has previously been suggested (Demere et al., 2003). Given the present North Atlantic distribution of the first diverging Phocina lineage, the grey seal (Figures 6-2, 6-4), an Arctic Atlantic origin is probable.

Approximately 1.5 Ma, the genera *Pusa* and *Phoca* diverged and both spread across the Arctic, with the ringed seal (*Pusa hispida*) retaining an Arctic distribution and *Phoca vitulina* entering both the northern Atlantic and Pacific. Between 1.3 and 0.4 Ma, the two *Phoca* species split, likely as a result of the retained (or re-attained) affinity of the spotted seal for ice and that of the harbour seal for land. Fossils resembling the extant *P. vitulina* and *P. largha* are not found until ~0.1-0.01 Ma (Miyazaki et al., 1994). This timing may instead represent the divergence time between the Atlantic and Pacific harbour seals, as spotted seals are only found in the Pacific in a neighbouring distribution to the Pacific harbour seals, but the harbour seal individuals used in this analysis were of Atlantic origin. The estimated molecular divergence between the Pacific spotted seal and the Atlantic harbour seals of ~1.3-0.4 Ma could coincide at the upper end with the first known occurrences of *Phoca*. *P. vitulina* is found in the eastern North Pacific ~1.3-1.2 Ma (Miyazaki et al., 1994) and the western North Atlantic ~1.64-0.79 Ma (Demere et al., 2003). Though the spotted seal and Pacific harbour seal may have diverged more recently, the time estimated here may be accurate, as spotted seals are differentiated from all harbour seals on several grounds and are the only subspecies of harbour seal to have been subsequently afforded species status (Shaughnessy and Fay, 1977). Subspecific relationships are beyond the scope of this paper and timing within *Phoca* cannot be definitive based on the present data.

Relationships within *Pusa* are not confidently resolved, but the ringed seal, *Pusa hispida*, and Baikal seal, *Pusa sibirica*, are recovered, but poorly supported, as sister (BPP=0.6685). Divergence between the three *Pusa* species is estimated to have occurred between 1.7 and 0.6 Ma (Table 6-3). These Phocina divergence times are generally congruent with those estimated by Sasaki et al. (2003) based on mt RFLPs, once their results are corrected to represent the mammalian molecular divergence (not substitution) rate of 2% per MY (see Palo and Vainola, 2006). The sister relationship recovered here (Figure 6-2) between the ringed and Baikal seals has been proposed numerous times (for review, see Palo and Vainola, 2006). This finding is consistent with the existence of an ancestral Arctic *Pusa* population, from which some individuals were isolated first in the Caspian sea, becoming *Pusa caspica*, and near simultaneously, an

invasion of Lake Baikal, which became *Pusa sibirica*. Davies (1958a) proposed a virtually simultaneous invasion of both the Caspian and Baikal by ringed seals, via the formation of a large lake due to the southern advance of Arctic glaciation, with the remaining ringed seals re-joining those populations that remained in the Arctic with glacial retreat. A Miocene *Paratethyan* origin of *Pusa* has also been proposed (Chapskii, 1955; McLaren, 1960a), but is not supported here, nor by other recent work (Demere et al., 2003; Palo and Vainola, 2006). Increasing the number of individuals used per species and the number of quickly-evolving DNA markers will be required to confidently resolve the relationships within *Pusa* (Chapter 5).

*Early Monachinae and the Monachini*

The tribe Monachini (monk seals) diverged from other monachines ~7.9 Ma, and the two extant monk seal species share their common ancestor ~6.2 Ma (Table 6-3, Figures 6-2, 6-3). The extinct *Pliophoca etrusca* has also been included in the Monachini and may represent the ancestor to the extant *Monachus* species (de Muizon, 1982). It is so morphologically similar to the extant *M. monachus* that it was originally described as the same species (see de Muizon, 1982). *P. etrusca* has been found in the western north Atlantic ~5.2-3.4 Ma and in the Mediterranean region ~3.4-1.64 Ma (Demere et al., 2003). Given these ages, several possibilities for the relationship between *Pliophoca* and *Monachus* exist. *P. etrusca* may be a sister lineage to one or both of the extant *Monachus* species that followed a similar geographic pattern, spreading across the Atlantic. If it does represent the ancestor to *Monachus*, it is possible that the divergence between extant *Monachus* species was actually in the western Atlantic, given the earlier occurrence there, as opposed to the more common hypothesis of a European divergence between monachine species (Hendey, 1972; Fyler et al., 2005). This could have involved some members of the ancestral *P. etrusca* stock remaining in the western Atlantic, eventually leading to the extinct Caribbean monk seal (*Monachus tropicalis*) and the Hawaiian monk seal (*M. schauinslandi*) and others dispersing to the Mediterranean to eventually become the Mediterranean monk seal (*M. monachus*). *Pristiphoca*, first found ~14-11 Ma in the Paratethys region, has also been included in the Monachini, though it is not likely to be the direct ancestor of *Monachus* and *Pliophoca* (de Muizon, 1982). The upper end of the 95% HPD for the divergence between Monachini and the remaining Monachinae is 9.285 Ma (Table 6-3), which is too recent for *Pristiphoca* to be a member of the Monachini. It is more likely that *Pristiphoca* may instead be one of the most recent stem Monachinae, along with the several species of *Monotherium* and

other genera found across the north Atlantic ~16.3-6.7 Ma. The Mediterranean origin of the extant monk seals is often based on the appearance of *Pristiphoca*, and sometimes on the presence of early monachines in several areas of Europe (Hendey, 1972). Several of the early Mediterranean or *Paratethyan* species, assigned as either monachines or phocines are regarded as having tendencies to the other subfamily (Repenning et al., 1979; de Muizon, 1982). This is consistent with early migrations of members of both subfamilies to various regions of Europe shortly after the divergence between subfamilies ~15 Ma and the retention of pleisiomorphies. Assuming a North American origin for the divergence between extant monk seals ~6 Ma, a second monachine dispersal including *P. etrusca* could have occurred. For a Mediterranean origin of the extant monk seals and inclusion of *Pristiphoca* in the Monachini, the estimated divergence of the Monachini ~7.8 Ma would have to have been erroneously underestimated by 2-3 million years and require an additional east to west dispersal for the ancestor of the Caribbean and Hawaiian monk seals (Hendey, 1972; Repenning et al., 1979; de Muizon, 1982). Based on the divergence times estimated here, the earlier occurrence of *Pliophoca* in North America than Europe, and the likely North American origin of the remaining monachine tribes (below), a North American origin for the Monachini is proposed (Figure 6-3, see also Arnason et al., 2006). Further investigation into the relationships between the early monachines and extant monk seals could yield insight into whether the divergence between extant monk seals did indeed occur in the western Atlantic.

*Miroungini*

The elephant seal tribe, Miroungini, diverged ~6.8 Ma from the Lobodontini, and the two *Mirounga* species share their common ancestor ~2.5 Ma. *Callophoca* may be the ancestor of the extant *Mirounga* (Ray, 1976; de Muizon, 1982), but it also shows similarities to *Monachus* (Repenning et al., 1979). Two species of *Callophoca* have been described, *C. obscura* and *C. ambigua*, and both are found in the eastern and western Atlantic (Ray, 1976). However, the latter species likely represents the male of the former species, and given the tendency for sexual dimorphism in the Miroungini, this has supported inclusion of *C. obscura* in the tribe (Ray, 1976; Repenning et al., 1979; de Muizon, 1982). *C. obscura* is found both in Belgium and along the eastern coast of the USA ~5.2-3.4 Ma (Demere et al., 2003). A North American origin for *Callophoca* has been proposed, with dispersal eastward either along the Gulf Stream in the northern Atlantic or through equatorial currents, to explain the European occurrence (de Muizon,

1982). The extant *Mirounga* evolved in the eastern Pacific, thus, their ancestor (potentially *Callophoca*) crossed through the Central American Seaway some time before it closed ~3 Ma, though the two species did not diverge until ~2.5 Ma. At this time, they likely began spreading northward and southward along the eastern Pacific until they reached their present distributions, where the northern elephant seal (*M. angustirostris*) resides along the northeastern Pacific coast and the southern elephant seal (*M. leonina*) in the Antarctic to sub-Antartic. The southern elephant seal had likely achieved its circumpolar distribution (Figure 6-4) during or prior to the Late Pleistocene (790-10 Ka), as fossils likely belonging to *Mirounga* are found in Australasia by this time (Demere et al., 2003). Remnant Atlantic populations would have gone extinct (Ray, 1976).

*Lobodontini*

The Lobodontini are comprised of the four Antarctic genera. The basal relationships are not confidently resolved (Chapter 5), and it is possible that either the Ross seal (*Ommatophoca*) or crabeater seal (*Lobodon*) may be the basal lineage or that the two may be sister taxa, as resolved here. There is no question, based on molecular evidence, that the leopard (*Hydrurga*) and Weddell (*Leptonychotes*) seals are a closely related sister group. However, an early morphological study placed the leopard and crabeater seals as sister and the Ross and Weddell seals as sister (de Muizon, 1982) and most researchers have followed this framework since. The association of the fossil taxa, *Piscophoca, Acrophoca,* and *Homiphoca* with the Lobodontini was proposed in the same study (de Muizon, 1982). Subsequent work did not ally these fossil taxa with specific extant lobodontine lineages, instead grouping the three fossil taxa together (Berta and Wyss, 1994). A European or American origin for the Lobodontini was originally discussed, based on an association with *Monotherium* species of both the eastern and western Atlantic ~10.4-6.7 Ma (de Muizon, 1982). *Piscophoca pacifica*, of Peru ~5.2-3.4 Ma, was proposed to be the descendent of this group of early monachines. The ~6.8 Ma origin of the tribe obtained here suggests that *Monotherium* is not part of the Lobodontini and a European origin is unlikely.

*Acrophoca* has been allied with the leopard seal (de Muizon, 1982) but the *Acrophoca* species of Peru are ~8-5 million years old (de Muizon and DeVries, 1985; Demere et al., 2003). This greatly predates the divergence of the leopard seal from the Weddell seal ~2.8 Ma. Given the discrepancy between molecular and morphological topologies, *Acrophoca* could be associated with the Weddell+leopard lineage, or any other Lobodontini. Since the Lobodontini are

estimated here to have split from the Miroungini ~6.7 Ma (5.692-7.898) and the two lobodontine sister groups (Figure 6-2) diverged ~5.4 Ma (4.381-6.495), the Peruvian *Acrophoca* species most likely represent stem lobodontines. The Pisco formation in Peru has also yielded several monachine species of uncertain origin, possibly Lobodontini (de Muizon and DeVries, 1985).

The last extinct genus to be associated with the Lobodontini is *Homiphoca capensis*, sometimes considered the potential ancestor of the crabeater seal (de Muizon, 1982). However, in their original description of *H. capensis*, Hendey and Repenning (1972) did not ally *Homiphoca* with a single extant species or lineage, but suggested that the reduction of the last upper postcanine tooth in *H. capensis* would "preclude the possibility of it being ancestral to *Hydrurga* or *Lobodon*". As the crabeater seal (*Lobodon*) diverged from other lobodontines ~5 Ma, the association between the crabeater seal and the 5.2-3.4 Ma South African *H. capensis* (Hendey and Repenning, 1972) is consistent.

Several possibilities exist for the association of the fossil taxa with extant Lobodontini. First, only those species resembling *Acrophoca* may be stem lobodontines and those resembling *Piscophoca* may represent another monachine lineage derived from *Monotherium aberratum* that migrated southward with the Lobodontini but left no living descendents. Second, *Piscophoca* may not be descendant of *Monotherium*, and may also be a stem lobodontine. Monachines have been known to exhibit very slow morphological change, as extant *Monachus* species resemble some of the earliest monachines (Hendey, 1972; Repenning and Ray, 1977; Wyss, 1988). Thus, it remains possible that the similarity between *Monachus* and *Piscophoca* may be pleisiomorphic. Third, while the early species of *Acrophoca* may be stem lobodontines, the later species, *A.. longirostris* of Peru (~5-3.5 Ma) and the ~5 Ma possible Chilean occurrence of *Acrophoca* (Walsh and Naish, 2002), may represent members of the leopard+Weddell lineage based on the similarity of *Acrophoca* to the leopard seal. *Piscophoca* are also found in Chile ~5 Ma, which could be associated with any lobodontine lineage or be unrelated. The possible crabeater lineage (*Homiphoca*) is found in South Africa ~5.2-3.4 Ma. Because the divergence times between the Ross seal, crabeater seal, and leopard+Weddell seals are highly overlapping with each other and with several fossil occurrences, none of the lobodontine fossil lineages can be excluded from any position within the Lobodontini, except for specific association with only the Weddell or the leopard seal, as they did not diverge from one another until ~2.8 Ma. Depending on the association of fossil and extant taxa, it is possible that up to three lineages of Lobodontini (crabeater, Ross, and leopard+Weddell ancestors) entered the Antarctic after a single, initial southward dispersal along the Pacific coast of South America.

The Atlantic to Pacific crossing of the ancestor of the Lobodontini through the Central American Seaway and subsequent southward movement ~6.8 Ma corresponds to the initiation of the rise of the Isthmus of Panama (McDougall, 1996). Between 6.7 and 6.4 Ma, the Caribbean warmed relative to the Pacific as deep water exchange ceased (McDougall, 1996) and the lobodontine ancestor may have maintained an association with cooler water. However, the timing and effects of the rise of the Isthmus of Panama are not fully understood or agreed upon (Molnar, 2008). The Lobodontini likely spread eastward around the Antarctic via the Antarctic Circumpolar Current (West wind drift), starting at least 3.4 Ma, based on the occurrence of *Homiphoca capensis* in South Africa ~5.2-3.4 Ma (Hendey and Repenning, 1972). While it is difficult to determine how many lobodontine lineages entered Antarctica, the extant taxa most probably obtained most of their specialized differences once they colonized this new environment (Hendey, 1972) and began to adapt to life with ice.

An interesting twist to Lobodontini evolution is the possible existence of *Homiphoca* along the central to south eastern US coast ~5.2-3.4 Ma (Ray, 1976). "A few bones" were first mentioned by Ray (1976) as resembling the South African *H. capensis* (then referred to as *Prionodelphis*). Though the genus appears well-described from South Africa, only recently have the North American samples been extensively studied. The North American material remained assigned to *Homiphoca*, though the assignment was stated to be quite tentative (Koretsky and Ray, 2008). Complex biogeographic hypotheses have been put forth to accommodate this unusual finding of a possible lobodontine (Demere et al., 2003). The referred species is found at the same time in South Africa, indicating long trans-oceanic dispersals would be required within a short time frame. It seems most parsimonious at present to accept the hypothesis of a single southward Lobodontini dispersal, with perhaps a wayward lineage retreating back northward and across the Central American Seaway to the Atlantic. Provided the North American material does not represent *H. capensis*, and represents other lobodontines or even other monachines, the simple single-dispersal theory holds. Further evaluation of fossil genera with extant genera, particularly with consideration for the molecular topology, will be required to be more confident in understanding lobodontine evolution.

**Conclusions**

Use of nuclear markers has presented both a new topology for the Phocidae and, in combination with multiple soft-bound fossil calibrations, new divergence time estimates.

Pinnipedia is estimated to have diverged from Musteloidea ~39 Ma and Phocidae from Otarioidea ~22 Ma, likely in the eastern Pacific.

The primarily Arctic subfamily, Phocinae, may have entered into Arctic conditions earlier than previously thought. The bearded seal diverged from the other phocines ~11.3 Ma, as the Arctic and Atlantic oceans became freely exchanging and perennial Arctic sea ice existed. The ancestor to the remaining Phocinae likely remained in the seas bordering between the Arctic and Atlantic oceans, with genera diverging between ~2 and 5 Ma with changing Arctic conditions. Speciations within *Phoca* and *Pusa* are not estimated to have occurred until ~1 Ma, when Arctic ringed seals invaded the Caspian Sea and Lake Baikal, probably from the north.

The two phocid subfamilies, Monachinae and Phocinae, split from one another ~14.7 Ma, probably on the eastern USA coast. If *Pliophoca etrusca* is assumed to be the ancestor of the extant *Monachus*, ancestral populations of *P. etrusca* may have split in the western north Atlantic ~6 Ma and some dispersed across the Atlantic to eventually become the Mediterranean monk seal and the remainder of the ancestral population remained, eventually becoming the now-extinct Caribbean monk seal and the extant Hawaiian monk seal. The two *Mirounga* species diverged from the Lobodontini ~ 6.8 Ma, crossing through the Central American Seaway and diverging from one another in the Pacific ~2.5 Ma. A single dispersal to the Antarctic along the western South American coast is hypothesized for the Lobodontini beginning ~6.8 Ma and eventually spreading around Antarctica. If the North American occurrence of *Homiphoca* ~5.2-3.4 Ma is validated and definitively associated with the Lobodontini, this single dispersal hypothesis will require change.

Most uncertainty in these hypotheses stems from questions regarding the association of fossil taxa with extant taxa. A comprehensive morphological examination of all fossil and extant genera now holds the greatest promise for understanding phocid biogeography.

**Table 6-1.** BEAST calibrations. Dates are represented in millions of years.

| Calibration | Fig. 6-1 Node # | Prior distribution | Offset | Mean | St. dev. | gamma shape / scale | Median age | 95% range |
|---|---|---|---|---|---|---|---|---|
| 1. Carnivora | 1 | lognormal | 40.0 | 1.5 | 1.0 | - | 44.482 | 40.865-63.22 |
| 2. Caniformia | 2 | normal | - | 38.0 | 1.5 | - | 40.00 | 35.07-44.93 |
| 3. Musteloidea-Pinnipedia | 3 | lognormal | 25.0 | 1.6 | 0.7 | - | 29.953 | 26.566-40.66 |
| 4. Otarioidea | 22 | gamma | 14.2 | - | - | 2.0 / 2.0 | 17.557 | 14.911-23.688 |
| 5. Phocidae | 5 | lognormal | 14.2 | 1.0 | 1.0 | - | 16.918 | 14.725-28.28 |
| 6. Phocina | 10 | lognormal | 0.5 | 1.5 | 1.0 | - | 4.982 | 1.365-23.72 |
| 7. Lobodontini-Miroungini | 17 | gamma | 5.2 | - | - | 3.5 / 1.5 | 9.959 | 6.826-15.75 |
| 8. Lobodontini | 19 | lognormal | 2.0 | 2.0 | 0.5 | - | 9.389 | 5.246-18.82 |
| **Alternate dates:** | | | | | | | | |
| 1b. Carnivora | 1 | normal | - | 65.0 | 5.0 | | 65.0 | 56.78-73.22 |

**Table 6-2.** Separate nuclear and mt BEAST divergence times. Nodes correspond to Figure 6-2.

| Node | Group | Nuclear only Mean divergence time (Ma) | 95% HPD range | | Mt only Mean divergence time (Ma) | 95% HPD range | |
|---|---|---|---|---|---|---|---|
| 1 | Carnivora (root) | 57.266 | 48.244 | 67.268 | 42.696 | 40.211 | 46.493 |
| 2 | Caniformia | 51.213 | 46.766 | 55.932 | 38.769 | 34.531 | 42.560 |
| 3 | Musteloidea-Pinnipedia | 39.977 | 35.536 | 44.234 | 33.030 | 28.570 | 37.110 |
| 4 | Pinnipedia | 22.052 | 19.694 | 24.640 | 24.646 | 21.052 | 28.311 |
| 5 | Phocidae | 14.705 | 14.263 | 15.289 | 16.273 | 14.375 | 18.616 |
| 6 | Phocinae | 11.223 | 8.052 | 13.945 | 12.968 | 10.351 | 15.629 |
| 7 | Cystophorini+Phocini | 4.033 | 2.884 | 5.222 | 9.086 | 6.935 | 11.367 |
| 8 | Phocini | 3.419 | 2.468 | 4.421 | 7.009 | 5.247 | 8.964 |
| 9 | Histriophocina | 2.727 | 1.824 | 3.641 | 5.065 | 3.359 | 6.827 |
| 10 | Phocina | 1.849 | 1.161 | 2.542 | 2.895 | 2.112 | 3.711 |
| - | *Halichoerus+Pusa* | 1.267 | 0.737 | 1.824 | - | - | - |
| 11 | *Phoca+Pusa* | - | | | 2.326 | 1.702 | 2.994 |
| 12 | *Phoca* | 0.669 | 0.249 | 1.172 | 1.337 | 0.863 | 1.849 |
| 13 | *Pusa* | 0.880 | 0.467 | 1.322 | - | - | - |
| - | *Phoca+Baikal+Ringed* | - | - | - | 2.157 | 1.567 | 2.782 |
| 14 | Baikal+Ringed | - | - | - | 1.890 | 1.281 | 2.547 |
| - | Caspian+Ringed | 0.655 | 0.321 | 1.031 | - | - | - |
| 15 | Monachinae | 7.093 | 5.983 | 8.247 | 11.921 | 9.783 | 14.201 |
| 16 | Monachini | 5.480 | 3.933 | 7.134 | 9.742 | 7.303 | 12.207 |
| 17 | Miroungini+Lobodontini | 6.342 | 5.484 | 7.354 | 9.897 | 8.021 | 11.932 |
| 18 | Miroungini | - | - | - | 2.684 | 1.591 | 3.841 |
| 19 | Lobodontini | 5.265 | 4.137 | 6.397 | 7.189 | 5.401 | 9.002 |
| - | Ross+Leopard+Weddell | - | - | - | 6.505 | 4.721 | 8.300 |
| 20 | Ross+Crabeater | 4.483 | 3.269 | 5.782 | - | - | - |
| 21 | Weddell+Leopard | 2.399 | 1.252 | 3.626 | 3.824 | 2.397 | 5.406 |
| 22 | Otarioidea | 15.026 | 14.213 | 16.119 | 18.067 | 15.207 | 20.974 |
| 23 | Otariidae | 1.831 | 0.932 | 2.868 | 7.001 | 4.666 | 9.476 |
| 24 | SAFS+NZFS | 0.281 | 0.037 | 0.590 | 1.793 | 1.012 | 2.734 |
| 25 | Mustelidae | 18.630 | 15.492 | 21.757 | 18.724 | 14.270 | 23.757 |
| 26 | Least weasel+Sea otter | 13.373 | 10.623 | 16.213 | 15.508 | 11.010 | 19.976 |
| 27 | Wolf+Arctic fox | 6.631 | 3.512 | 7.390 | 16.507 | 10.341 | 23.128 |
| 28 | Cat+Lynx | 5.329 | 4.410 | 8.992 | 9.877 | 5.800 | 14.666 |

**Table 6-3.** BEAST nuclear + mt divergence times. Nodes correspond to Figure 6-2.

| Node | Group | Mean divergence time (Ma) | 95% HPD range | |
|---|---|---|---|---|
| 1 | Carnivora (root) | 54.395 | 46.458 | 63.895 |
| 2 | Caniformia | 50.313 | 45.462 | 55.218 |
| 3 | Musteloidea-Pinnipedia | 39.412 | 34.928 | 44.176 |
| 4 | Pinnipedia | 22.149 | 19.769 | 24.739 |
| 5 | Phocidae | 14.683 | 14.266 | 15.243 |
| 6 | Phocinae | 11.341 | 9.133 | 13.430 |
| 7 | Cystophorini+Phocini | 5.371 | 3.987 | 6.744 |
| 8 | Phocini | 4.309 | 3.174 | 5.474 |
| 9 | Histriophocina | 3.307 | 2.306 | 4.456 |
| 10 | Phocina | 1.947 | 1.387 | 2.524 |
| 11 | *Phoca+Pusa* | 1.543 | 1.080 | 2.010 |
| 12 | *Phoca* | 0.789 | 0.418 | 1.267 |
| 13 | *Pusa* | 1.218 | 0.766 | 1.665 |
| 14 | Baikal+Ringed | 1.094 | 0.647 | 1.542 |
| 15 | Monachinae | 7.880 | 6.501 | 9.285 |
| 16 | Monachini | 6.219 | 4.656 | 7.745 |
| 17 | Miroungini+Lobodontini | 6.793 | 5.692 | 7.898 |
| 18 | Miroungini | 2.522 | 0.932 | 3.789 |
| 19 | Lobodontini | 5.424 | 4.381 | 6.495 |
| 20 | Ross+Crabeater | 4.899 | 3.810 | 6.017 |
| 21 | Weddell+Leopard | 2.813 | 1.778 | 3.834 |
| 22 | Otarioidea | 15.095 | 14.226 | 16.267 |
| 23 | Otariidae | 2.636 | 1.729 | 3.758 |
| 24 | S.American+New Zealand FS | 0.679 | 0.357 | 1.093 |
| 25 | Mustelidae | 18.814 | 15.814 | 21.894 |
| 26 | Least weasel+Sea otter | 14.237 | 11.504 | 17.101 |
| 27 | Wolf+Arctic fox | 7.677 | 5.222 | 10.416 |
| 28 | Cat+Lynx | 6.281 | 4.275 | 8.410 |

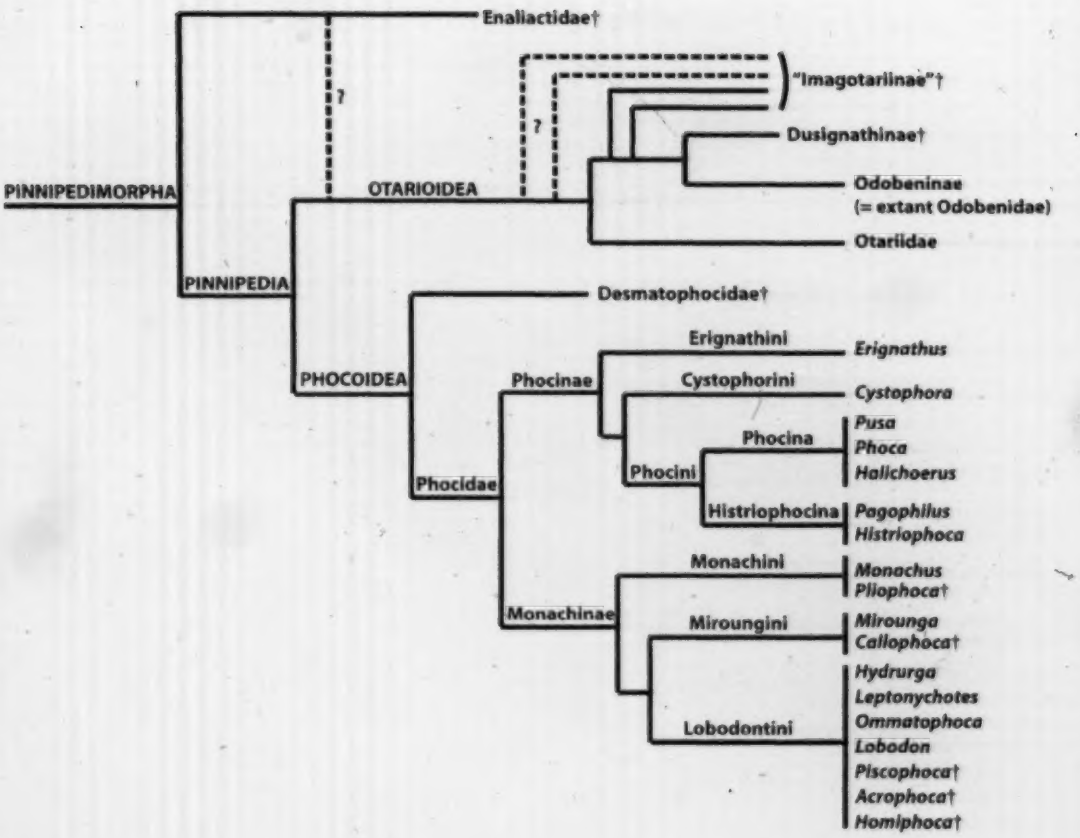


**Figure 6-1.** Overview of pinniped relationships of extant and some extinct taxa. Fossil taxa discussed in text are listed with their putative subfamily. The relationships between extinct and extant subfamilies and families are a composite of the results of the most compatible molecular and morphological studies. Dashed lines represent possible alternate branching structure.

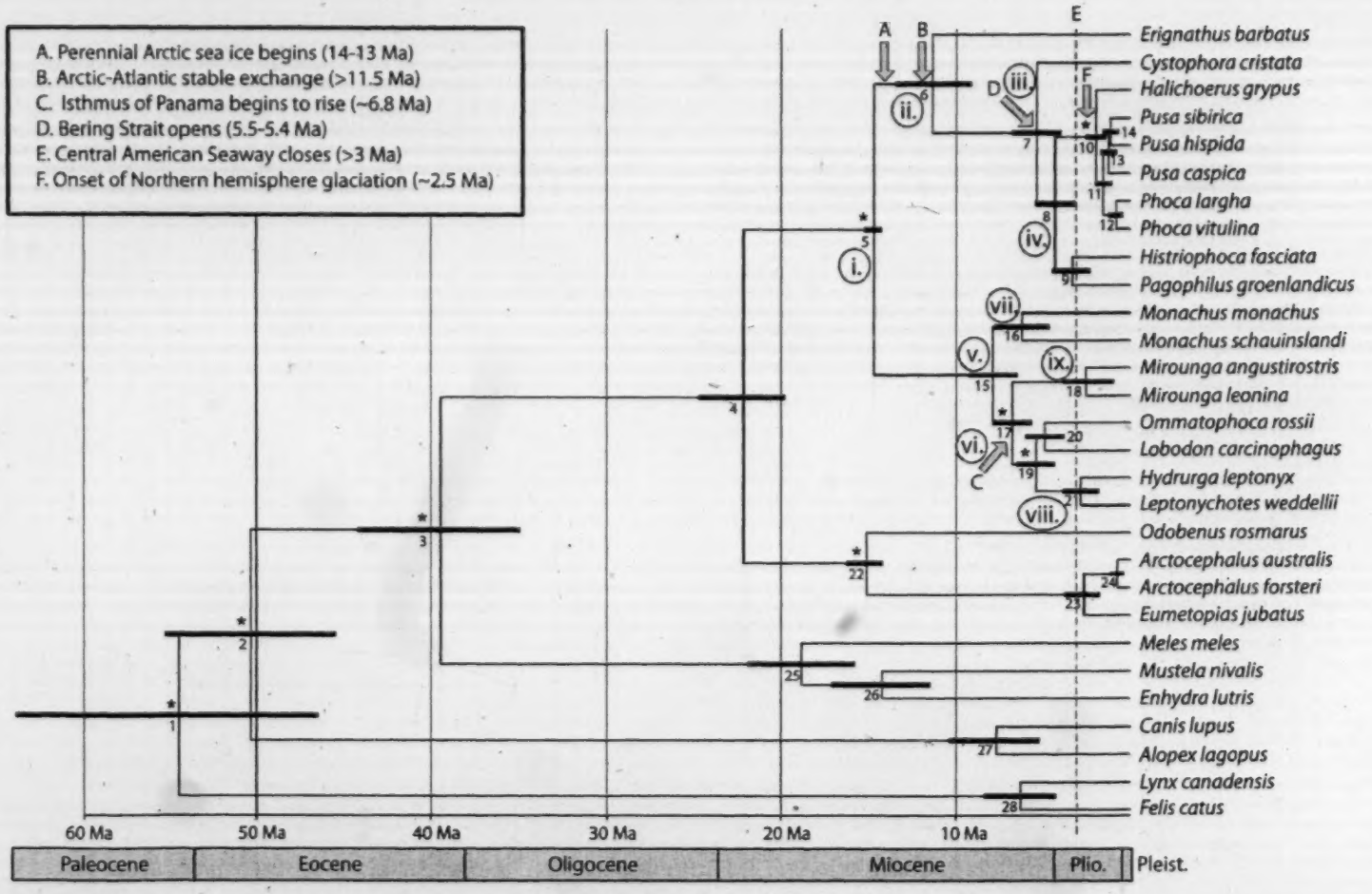


**Figure 6-2.** Combined nuclear + mt BEAST phylogeny. Bars represent the 95% HPD interval. Node numbers correspond to Table 6-3 for divergence time estimates. Stars represent fossil calibrated nodes (Table 6-1) and Roman numerals indicate events in Figure 6-3.

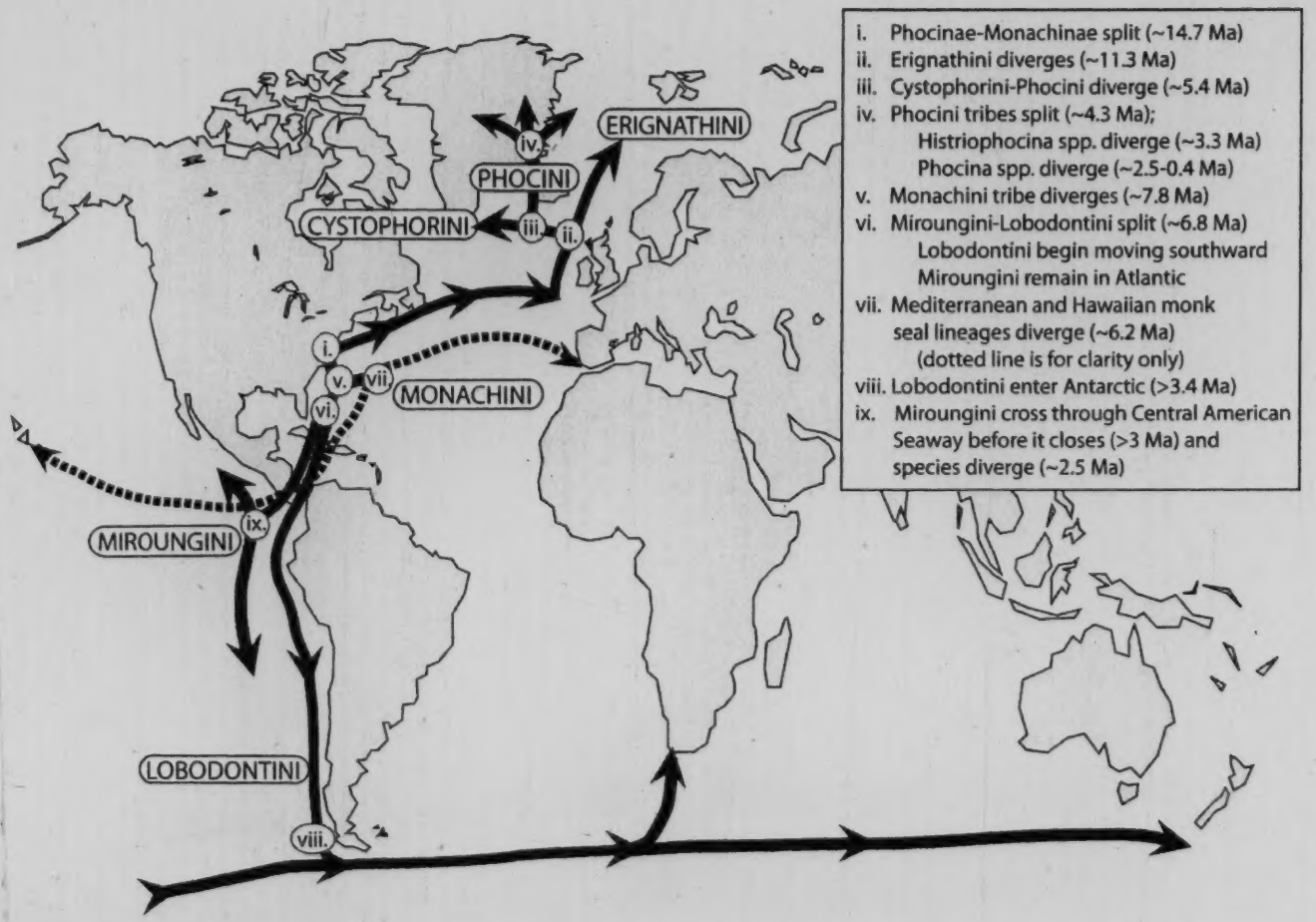


**Figure 6-3.** Hypothesized historical movements of phocid lineages. Divergence times with HPD ranges are in Table 6-3 and the complete phylogeny is shown in Figure 6-2.

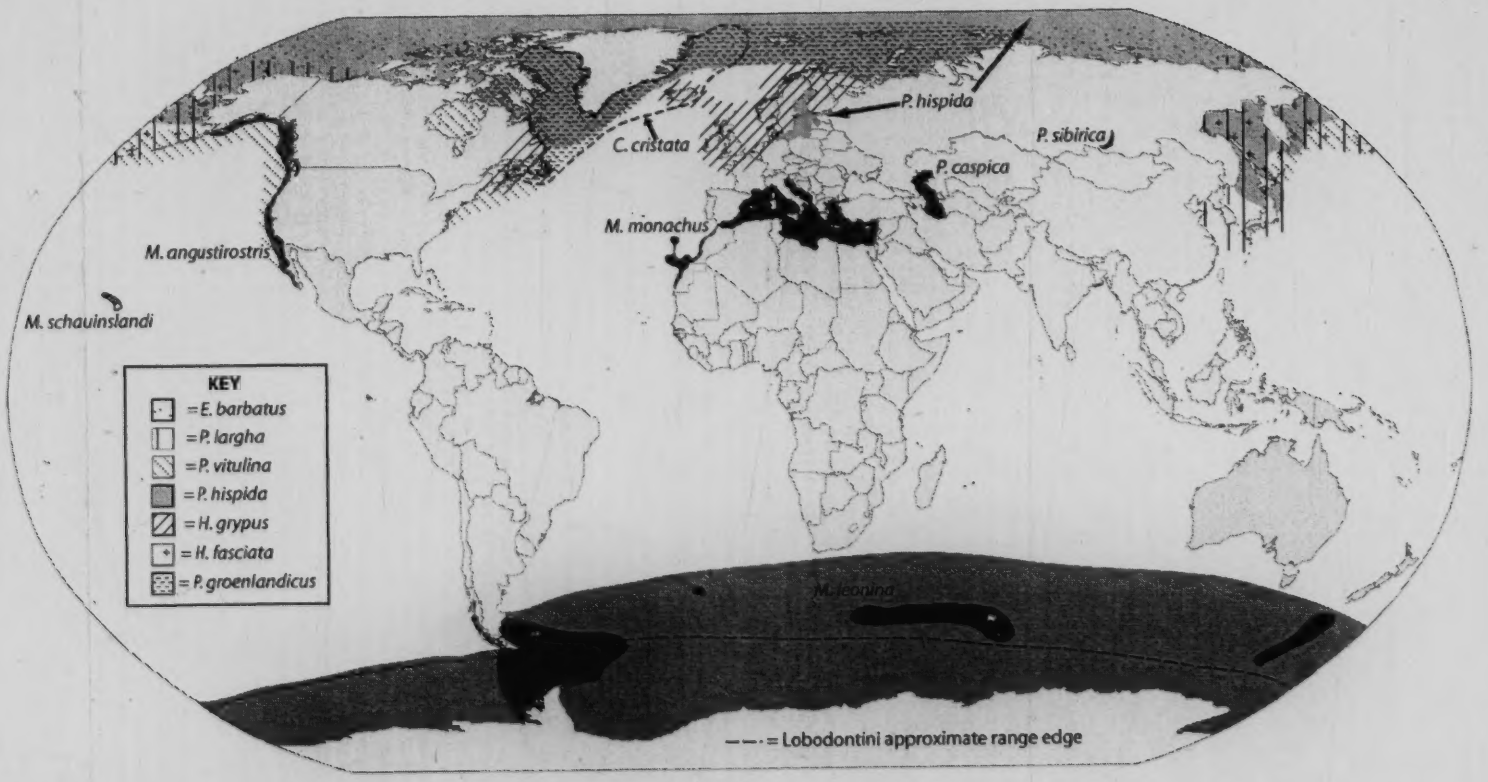


**Figure 6-4.** Present distribution of phocids. The approximate range of each extant phocid is labelled by shading or, in the case of most Phocinae (whose ranges are overlapping), by a pattern indicated in the key. For both *Mirounga* species, the breeding area is indicated in dark grey and the range of non-breeding individuals in surrounding medium grey. Range information was obtained from King (1983).

**References**


Arnason, U., 1974. Comparative chromosome studies of Pinnipedia. Hereditas 76, 179-226.

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., Lehman, N., Petrov, E.A., Vainola, R., 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol.Phylogenet.Evol. 41, 345-354.

Barnes, L.G., 1987. An early Miocene pinniped of the genus *Desmatophoca* (Mammalia: Otariidae) from Washington. Contributions in Science. Natural History Museum of Los Angeles County, Los Angeles, CA, p. 20.

Baskin, J.A., 1998. Mustelidae. In: Janis, C.M., Scott, K.M., Jacobs, L.J. (Eds.), Evolution of Tertiary mammals of North America. Cambridge University Press, New York, NY, pp. 152-173.

Benton, M.J., Donoghue, P.C.J., 2007. Paleontological evidence to date the tree of life. Mol.Biol.Evol. 24, 26-53.

Berta, A., 1991. New *Enaliarctos** (Pinnipedimorpha) from the Oligocene and Miocene of Oregon and the role of "Enaliarctids" in pinniped phylogeny. Smith.Contrib.Paleobio. 69, 1-33.

Berta, A., Ray, C.E., Wyss, A.R., 1989. Skeleton of the oldest known pinniped, *Enaliarctos mealsi*. Science 244, 60-62.

Berta, A., Sumich, J.L., Kovacs, K.M., 2006. Marine mammals: evolutionary biology. Elsevier Academic Press.

Berta, A., Wyss, A.R., 1994. Pinniped phylogeny. Proc.San Diego Soc.Natl.Hist. 29, 33-56.

Bininda-Emonds, O.R.P., Russell, A.P., 1996. A morphological perspective on the phylogenetic relationships of the extant phocid seals (Mammalia: Carnivora: Phocidae). Bonner Zooligische Monographien, Nr. 41. 256 pp.

Burns, J.J., Fay, F.H., 1970. Comparative morphology of the skull of the Ribbon seal, *HistrioHistriophoca fasciata*, with remarks on systematics of Phocidae. J.Zool.Lond. 161, 363-394.

Chapskii, K.K., 1955. Contribution to the problem of the history of development of the Caspian and Baikal seals. Trudy Zool. Inst. Akad. Nauk SSSR 17, 200-216.

Cozzuol, M.A., 2001. A "northern" seal from the Miocene of Argentina: implications for phocid phylogeny and biogeography. J.Vert.Paleo. 21, 415-421.

Darby, D.A., 2008. Arctic perennial ice cover over the last 14 million years. Paleoceanography 23, PA1S07.

Davies, J.L., 1958. Pleistocene geography and the distribution of northern pinnipeds. Ecology 39, 97-113.

Davis, C.S., Delisle, I., Stirling, I., Siniff, D.B., Strobeck, C., 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. Mol.Phylogenet.Evol. 33, 363-377.

Davis, C.S., Stirling, I., Strobeck, C., Coltman, D.W., 2008. Population structure of ice-breeding seals. Mol.Ecol. 17, 3078-3094.

de Muizon, C., 1982. Phocid phylogeny and dispersal. Ann.S.Afr.Mus. 89, 175-213.

de Muizon, C., DeVries, T.J., 1985. Geology and paleontology of late Cenozoic marine deposits in the Sacaco area (Peru). Geologische Rundschau 74, 547-563.

de Muizon, C., Hendey, Q.B., 1980. Late Tertiary seals of the South Atlantic ocean. Ann.S.Afr.Mus. 82, 91-128 *not seen by this author.

Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. Mol.Phylogenet.Evol. 37, 192-201.

Demere, T.A., 1994. The family Odobenidae: A phylogenetic analysis of fossil and living taxa. Proc.San Diego Soc.Natl.Hist. 29, 99-123.

Demere, T.A., Berta, A., 2001. A reevaluation of *Proneotherium repenningi* from the Miocene Astoria formation of Oregon and its position as a basal odobenid (Pinnipedia: Mammalia). J.Vert.Paleo. 21, 279-310.

Demere, T.A., Berta, A., Adam, P.J., 2003. Pinnipedimorph evolutionary biology. Bull.Am.Mus.Nat.Hist. 13, 32-76.

Dragoo, J.W., Honeycutt, R.L., 1997. Systematics of Mustelid-like carnivores. J.Mammal 78, 426-443.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenies and dating with confidence. PLOS Biology 4, e88.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol.Biol. 7, 214.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J.Mol.Evol. 17, 368-376.

Finarelli, J.A., 2008. A total evidence phylogeny of the Arctoidea (Carnivora: Mammalia): relationships among basal taxa. J.Mamm.Evol. in press.

Flower, W.H., Lyddekker, R., 1891. An introduction to the study of mammals living and extinct. Adam & Charles Black, London.

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M.A., 2005. Molecular phylogeny of the Carnivora (Mammalia): Assessing the impact of increased sampling on resolving enigmatic relationships. Syst.Biol. 54, 317-337.

Flynn, J.J., Nedbal, M.A., 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs incompatibility among multiple data sets. Mol.Phylogenet.Evol.9, 414-426.

Flynn, J.J., Wesley-Hunt, G.D., 2005. Carnivora. In: Rose, K.D., Archibald, J.D. (Eds.), The rise of placental mammals: origins and relationships of the major extant clades. The John Hopkins University Press, Baltimore, Maryland, pp. 175-198.

Fulton, T.L., Strobeck, C., 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. Mol.Phylogenet.Evol. 41, 165-181.

Fyler, C.A., Reeder, T.W., Berta, A., Antonelis, G., Aguilar, A., Androukaki, E., 2005. Historical biogeography and phylogeny of monachine seals (Pinnipedia: Phocidae) based on mitochondrial and nuclear DNA data. J.Biogeogr. 32, 1267-1279.

Gladenkov, A.Y., Oleinik, A.E., Marincovich Jr., L., Barinov, K.B., 2002. A refined age for the earliest opening of Bering Strait. Palaeogeogr. Palaeoclim. Palaeoecol. 183, 321-328.

Gregory, W.K., Hellman, M., 1939. On the evolution and major classification of the civets (Viverridae) and allied fossil and recent Carnivora: Phylogenetic study of the skull and dentition. Proc.Am.Phil.Soc. 81, 309-392.

Haley, B.A., Frank, M., Speilhagen, R.F., Fietzke, J., 2008. Radiogenic isotope record of Arctic Ocean circulation and weathering inputs of the past 15 million years. Paleoceanography 23, PA1S13.

Harington, C.R., 2008. The evolution of Arctic marine mammals. Ecol.App. 18, S23-S40.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J.Mol.Evol. 22, 160-174.

Hendey, Q.B., 1972. The evolution and dispersal of the Monachinae (Mammalia: Pinnipedia). Ann.S.Afr.Mus. 59, 99-113.

Hendey, Q.B., Repenning, C.A., 1972. A Pliocene phocid from South Africa. Ann.S.Afr.Mus. 59, 71-98.

Higdon, J.W., Bininda-Emonds, O.R.P., Beck, R.M.D., H., F.S., 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a mutigene dataset. BMC Evol.Biol. 7, 216.

Ho, S.Y.W., 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology 38, 409-414.

Hugall, A.F., Foster, R., Lee, M.S.Y., 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Syst.Biol. 56, 543-563.

Illiger, J.C.W., 1811. Prodomus systematics Mammalium et Avium. C. Salfeld, Berlin, Germany.

Kimura, M., 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J.Mol.Evol. 16, 111-120.

King, J.E., 1966. Relationships of the Hooded and Elephant seals (genera *Cystophora* and *Mirounga*). J.Zool. 148, 385-398.

King, J.E., 1983. Seals of the world. British Museum (Natural History) and Oxford University Press, London and Oxford.

Kohno, N., 1994. A new Miocene pinniped in the genus *Prototaria* (Carnivora: Odobenidae) from the Moniwa Formation, Miyagi, Japan. J.Vert.Paleo. 14, 414-426.

Kohno, N., 2006. A new Miocene odobenid (Mammalia: Carnivora) from Hokkaido, Japan, and its implications for odobenid phylogeny. J.Vert.Paleo. 26, 411-421.

Kohno, N., Barnes, L.G., Hirota, K., 1994. Miocene fossil pinnipeds of the genera *Prototaria* and *Neotherium* (Carnivora; Otariidae; Imagotariinae) in the North Pacific Ocean: Evolution, relationships and distribution. The Island Arc 3, 285-308.

Koretsky, I.A., 2001. Morphology and systematics of Miocene Phocinae (Mammalia: Carnivora) from Paratethys and the North Atlantic region. Geologica Hungarica. Ser.Palaeo.Fasc. 54, 1-109.

Koretsky, I.A., Holec, P., 2002. A primitive seal (Mammalia: Phocidae) from the Early Middle Miocene of Central Paratethys. In: Emry, R.J. (Ed.), Cenozoic mammals of land and sea: Tributes to the career of Clayton E. Ray. Smithsonian Institute Press, Washington, D.C., pp. 163-178.

Koretsky, I.A., Ray, C.E., 2008. Phocidae of the Pliocene of eastern USA. In: Ray, C.E., Bohaska, D.J., Koretsky, I.A., Ward, L.W., Barnes, L.G. (Eds.), Geology and Paleontology of the Lee Creek Mine, North Carolina IV. Virginia Museum of Natural History, pp. 81-140.

Koretsky, I.A., Sanders, A.E., 2002. Paleontology of the Late Oligocene Ashley and Chandler Bridge formations of South Carolina, 1: Paleogene pinniped remains; the oldest known seal (Carnivora; Phocidae). In: Emry, R.J. (Ed.), Cenozoic mammals of land and sea: Tributes to the career of Clayton E. Ray. Smithsonian Institution Press, Washington, D.C., pp. 179-183.

Kovacs, K.M., Lydersen, C., Gjertz, I., 1996. Birth-site characteristics and prenatal molting in bearded seals (*Erignathus barbatus*). J.Mammal. 77, 1085-1091.

Krylov, A.A., Andreeva, I.A., Vogt, C., Backman, J., Krupskaya, V.V., Frikurov, G.E., Moran, K., Shoji, H., 2008. A shift in heavy and clay mineral provenance indicates a middle Miocene onset of perennial sea ice cover in the Arctic Ocean. Paleoceanography 23, PA1S06.

Ledje, C., Arnason, U., 1996. Phylogenetic relationships within caniform carnivores based on analyses of the mitochondrial 12S rRNA gene. J.Mol.Evol. 43, 641-649.

Lunt, D.J., Valdes, P.J., Haywood, A., Rutt, I.C., 2008. Closure of the Panama Seaway during the Pliocene: implications for climate and Northern Hemisphere glaciation. Clim.Dynam. 30, 1-18.

Lydersen, C., Kovacs, K.M., 1999. Behaviour and energetics of ice-breeding, North Atlantic phocid seals during the lactation period. Mar.Ecol.Prog.Ser. 187, 265-281.

Lyle, M., Barron, J., Bralower, T.J., Huber, M., Lyle, A.O., Ravelo, A.C., Rea, D.K., Wilson, P.A., 2008. Pacific Ocean and Cenozoic evolution of climate. Rev.Geophysics 46, 1-47.

McDougall, K., 1996. Benthic foraminiferal response to the emergence of the Isthmus of Panama and coincident paleoceanographic changes. Mar.Micropaleo. 28, 133-169.

McLaren, I.A., 1960a. On the origin of the Caspian and Baikal seals and the paleoclimatological implication. Am.J.Sci. 258, 47-65.

McLaren, I.A., 1960b. Are the Pinnipedia biphyletic? Syst.Zool. 9, 18-28.

Mitchell, E., 1968. The Mio-Pliocene Pinniped *Imagotaria*. J.Fish.Res.Board Canada 25, 1843-1900.

Mitchell, E., Tedford, R.H., 1973. The Enaliarctinae A new group of extinct aquatic Carnivora and a consideration of the origin of the Otariidae. Bull.Am.Mus.Nat.Hist. 151, 201-284.

Mivart, S.G., 1885. Notes on the Pinnipedia. Proc. Zool. Soc. London 1885, 484-484-500.

Miyazaki, S., Horikawa, H., Kohno, N., Hirota, K., Kimura, M., Hasegawa, Y., Tomida, Y., Barnes, L.G., Ray, C.E., 1994. Summary of the fossil record of pinnipeds of Japan, and comparisons with that from the eastern North Pacific. The Island Arc 3, 361-372.

Molnar, P., 2008. Closing of the Central American Seaway and the Ice Age: A critical review. Paleoceanography 23, PA2201.

Munthe, K., 1998. Canidae. In: Janis, C.M., Scott, K.M., Jacobs, L.J. (Eds.), Evolution of Tertiary mammals of North America. Cambridge University Press, Cambridge, UK, pp. 124-143.

Nylander, J.A.A., 2004. MrModeltest. Version 2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Palo, J.U., Vainola, R., 2006. The enigma of the landlocked Baikal and Caspian seals addressed through phylogeny of phocine mitochondrial sequences. Biol.J.Linn.Soc. 88, 61-72.

Posada, D., Crandall, K.A., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Rambaut, A., Drummond, A.J., 2007. Tracer. v1.4. http://tree.bio.ed.ac.uk/software/tracer/

Ray, C.E., 1976. Geography of phocid evolution. Syst.Zool. 25, 391-406.

Repenning, C.A., Ray, C.E., 1977. The origin of the Hawaiian monk seal. Proceedings of the Biological Society of Washington 89, 667-688.

Repenning, C.A., Ray, C.E., Grigorescu, D., 1979. Pinniped biogeography. In: Boucot, A.J. (Ed.), Historical biogeography, plate tectonics, and thee changing environment: proceedings of the thirty-seventh annual biology colloquium and selected papers. Oregon State University Press, pp. 357-369.

Sasaki, H., Numachi, K., Grachev, M.A., 2003. The origin and genetic relationships of the Baikal seal, *Pusa sibirica*, by restriction analysis of mitochondrial DNA. Zool.Sci. 20, 1417-1422.

Sato, J.J., Wolsan, M., Suzuki, H., Hosoda, T., Yamaguchi, Y., Hiyama, K., Kobayashi, M., Minami, S., 2006. Evidence from nuclear DNA sequences sheds light on the phylogenetic relationships of pinnipedia: single origin with affinity to musteloidea. Zool.Sci. 23, 125-146.

Shaughnessy, P.D., Fay, F.H., 1977. A review of the taxonomy and nomenclature of North Pacific Harbour seals. J.Zool.Lond. 182.

Simpson, G.G., 1945. The principles of classification and a classification of mammals. Bull.Am.Mus.Nat.Hist. 85, 1-350.

Stirling, I., 1977. Adaptations of Weddell and ringed seals to exploit the polar fast ice habitat in the absence or presence of surface predators. In: Llano, G.A. (Ed.), Adaptations within Antarctic ecosystems, pp. 741-748.

Stirling, I., 1983. The evolution of mating systems in pinnipeds. In: Eisenberg, J.F., Kleiman, D. (Eds.), Recent advances in the study of mammalian behaviour. Special Publication No. 7 of the American Society of Mammalogists, pp. 489-527.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sunderland, Massachusetts.

Walsh, S., Naish, D., 2002. Fossil seals from Late Neogene deposits in South America: a new pinniped (Carnivora, Mammalia) assemblage from Chile. Palaeontology 45, 821-842.

Wang, X., 1994. Phylogenetic systematics of the Hesperocyoninae (Carnivora: Canidae). Bull.Am.Mus.Nat.Hist. 221, 1-207.

Wang, X., McKenna, M.C., Dashzeveg, D., 2005. *Amphicticeps* and *Amphicynodon* (Arctoidea, Carnivora) from Hsanda Gol formation, central Mongolia and phylogeny of basal arctoids with comments on zoogeography. Am.Mus.Novit. American Museum of Natural History, New York, NY, p. 57.

Welch, J.J., Bromham, L., 2005. Molecular dating when rates vary. Trends Ecol.Evol. 20, 320-327.

Wesley-Hunt, G.D., Flynn, J.J., 2005. Phylogeny of the Carnivora: basal relationships among the Carnivoramorphans, and assessment of the position of 'Miacoidea' relative to Carnivora. J.Syst.Palaeo. 3, 1-28.

Wyss, A.R., 1987. The walrus auditory region and the monophyly of the pinnipeds. Am.Mus.Novit. 2871, 1-31.

Wyss, A.R., 1988. On "retrogression" in the evolution of the Phocinae and phylogenetic affinities of the monk seals. Am.Mus.Novit. 2924, 1-38.

Wyss, A.R., Flynn, J.J., 1993. A phylogenetic analysis and definition of the Carnivora. In: Szalay, F., Novacek, M., McKenna, M. (Eds.), Mammal Phylogeny: Placentals. Springer-Verlag, New York, pp. 32-52.

Yang, Z., Rannala, B., 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol.Biol.Evol. 23, 212-226.

Zharkikh, A., 1994. Estimation of evolutionary distances between nucleotide sequences. J.Mol.Evol. 39.

# Chapter 7

## General discussion and conclusions

### Resolution for Carnivora: many genes, many analyses, many answers

The use of multiple loci, particularly the application of multiple independent nuclear loci, in multiple analyses has provided substantial resolution to the arctoid carnivore phylogeny. Tribal, subfamilial, and familial structure within Phocidae (true seals) were strongly supported by all analyses and datasets (Chapter 5), as was pinniped monophyly (Chapters 2, 5, 6), and the sister relationship of Pinnipedia with Musteloidea (Chapter 2). Within Musteloidea, the basal position of the skunk family, Mephitidae, was proposed (Chapter 2) and the difficulty in resolving the relationships between Mephitidae, Ailuridae, and Procyonidae+Mustelidae was illustrated to be the result of a rapid radiation, where too little time to accumulate synapomorphies, data biases, and deep coalescent events combined to cause locus and analysis incongruence (Chapter 3). However, the concatenation of nuclear markers provided a strongly supported (via bootstrap and statistical testing) topology, with Mephitidae branching first, then Ailuridae, then Procyonidae+Mustelidae. Within musteloid families, subfamily structures for Procyonidae and Mustelidae were clarified (Chapter 2,4), as the previous procyonid subfamilies were shown to be based upon convergent morphological characters (Chapter 4).

For the sections of the phylogeny that remain problematic, future solutions have been outlined through identification of the factors leading to difficulty in resolution. For example, excluding DNA partitions that exhibited compositional bias across taxa and including multiple individuals per species allowed the first molecular estimate of the northern seal *Phoca-Pusa-Halichoerus* complex that was fairly consistent with morphology (Chapter 5). Expansion of this work to include many more individuals per species within a coalescent framework should provide greater insight into the relationships between *Pusa* species. If the *Pusa* relationships can be definitively resolved, insight into the paleogeography of Eurasia may be possible, as Arctic connections may be identified with the Caspian Sea and Lake Baikal, explaining the occurrence of the land-locked seals in these water bodies. Increasing the number of individuals per species and accounting for data biases will be useful in resolving many of the rapidly radiated Otariidae (fur seal and sea lions) and Ursidae (bears) relationships.

Using multiple nuclear loci and multiple analyses were integral to resolving most of these phylogenetic problems. Multiple independent nuclear loci can provide a more comprehensive

view of species evolution, sometimes resolving relationships that mitochondrial DNA cannot (i.e. Ursidae-Pinnipedia-Musteloidea). Such a comprehensive view is important in resolving rapid radiations where phylogenetic artifacts, deep coalescent events, and a lack of phylogenetic signal make the species tree unclear. If only one gene tree had been examined for basal musteloids, the inferred species tree would have differed dramatically based on the selected gene (Chapter 3). Analyzing loci separately, as for the case of the red panda (Chapter 3) can illustrate discordance among genes that may be hidden when loci are concatenated. In two cases, not only were loci discordant for relationships under investigation, but some loci were identified as incongruent with all other loci and multiple other independent lines of evidence. Excluding such incongruent loci is an important factor for obtaining accurate phylogenies in any analysis. But as the field of phylogenetics moves into the realm of phylogenomics at a rapid pace, recognized discordance and incongruence will increase, as species trees will always be inferred from multiple loci, chromosomes, or even complete genomes. Thus, it is important to continue to deconstruct the concatenated signal into its possibly incongruent parts, be those individual loci, linkage units, or chromosomes. Recent advances in species tree construction from gene trees have been made (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007; Liu and Pearl, 2007), though most reconstruction algorithms are still susceptible to any data biases that violate model assumptions such as base compositional bias, and mutation rate differences across lineages or through time. Though these biases are being increasingly better identified and modeled, generally in a computer-simulated framework (Galtier and Gouy, 1995; Jermiin et al., 2004; Lockhart et al., 2005; Philippe et al., 2005; Blanquart and Lartillot, 2006; Ruano-Rubio and Fares, 2007; Zhou et al., 2007), it is important to test and apply these methods to empirical data.

Data biases such as incomplete taxon or locus sampling (Chapter 2), base composition bias across taxa (Chapters 3, 5), locus incongruence (Chapters 3, 5), and rate heterogeneity across sites (Chapters 2-6) and lineages (Chapter 3) have been identified throughout this thesis. The results of the empirical evaluation of many phylogenetic reconstruction methods when such biases exist provide both good and bad news for the performance of the commonly applied algorithms. It appears that most common reconstruction methods are equally robust/susceptible to the effects of missing data, as it is the signal from the data in hand that is the most important factor. With ever-increasing DNA sequence information in public databases and the incomplete nature of some of the genome-wide sequencing strategies, such as ESTs, the amount of data that is missing across taxa or loci will only increase over the short term. In the case of the position of the red panda, present algorithms did not appear able to "find" the phylogenetic signal for the

noise, particularly minimum evolution and maximum parsimony. Reconstruction was greatly improved through the use of maximum likelihood in a framework that allows an independent set of parameters to be estimated for each locus, as opposed to forcing a single set of parameters to explain the entirety of a heterogenous dataset. When too few individuals per species were used, all of the algorithms were misled to a certain extent by data biases and incomplete lineage sorting within Phocina (northern seal subtribe of Phocini). This study (Chapter 5) also illustrated the effect of different partitioning strategies on mitochondrial DNA results, as preferred methods of partitioning the DNA sequence resulted in increased support, and different strategies recovered different topologies. Perhaps the most important conclusion from performing multiple analyses is simply the importance of using of multiple analysis methods to more easily allow identification of problematic areas within a phylogeny and instigate further investigation into the causes of discordance between methods.

**Themes in carnivore evolution**

Like many groups of organisms, Carnivora contains many rapid radiations. Some have been easier to resolve than others and a few remain unresolved. Unraveling these radiations provides a strong framework within which to determine the factors involved in rapid radiations. By resolving the arctoid trichotomy and the musteloid family relationships, the higher-level structure of the Caniformia is now better understood. As was applied for the seals in Chapter 6 and recently for mustelids (Koepfli et al., 2008), incorporation of fossil taxa into molecular dating techniques allows biogeographic inference. Biogeographic study is now possible for the deeper divergences across Caniformia and many of the more recent divergences within families, obtaining insight into the locations that divergences took place and the ecological and environmental conditions under which rapid radiations occur. Linking phylogeny and the environment allow more general inferences about the factors that lead to diversification and extinction. When phylogenies across large groups such as the Caniformia become available, inferences about the tempo and mode of evolution is no longer restricted to small sections of the phylogeny, but can be compared and contrasted across multiple families and superfamilies. Rapid radiations are observed at every timescale across the Tree of Life, thus understanding the factors behind such radiations across groups of organisms can provide has far-reaching implications. Using multiple markers and analysis methods has provided resolution for many

such radiations within Arctoidea; the next step is to use this framework to understand why species radiations occur.

Another phenomenon observed across carnivoran groups is morphological convergence. Molecular analyses are generally more susceptible to problems of convergence, but are often easier to identify and account for with more complex nucleotide evolution models. Within Arctoidea, procyonid genera exhibit high levels of convergence in their adaptations to highly similar environments (Chapter 4, Koepfli et al., 2007). The Antarctic seals, *Hydrurga* (leopard seal) and *Lobodon* (crabeater seal), have likely converged on similar dentition, specializing on krill as a food source. Both cases were identified through the use of molecular markers. The converse is also observed within Phocidae, where convergence in nucleotide composition combined with other confounding factors resulted in strong support for the paraphyly of the morphologically-conserved genus *Pusa* by the inclusion of the highly divergent *Halichoerus*. In this case, morphology clearly defined genera and it was only by identifying molecular convergence that the two marker types could be reconciled. Convergence in both marker types is likely much more prevalent that commonly recognized, as DNA regions under strong selection or those that are highly mutable are often excluded from phylogenetic analyses, as are morphological characters that are thought to be adaptive. A strong molecular framework, combined with strong morphological work, can provide insight into morphological adaptations and disentangle homoplasy from plesiomorphic or synapomorphic characters.

**Future directions**

Though most have been answered, some phylogenetic questions remain within Arctoidea. Ursid phylogeny is improving, but conflict remains between recent studies (Yu et al., 2007; Nakagome et al., 2008; Pages et al., 2008). Otariidae phylogeny has proven difficult and also remains poorly resolved (Wynen et al., 2001; Higdon et al., 2007), as do the species relationships within the northern seal subtribe Phocina (Chapter 5). Each of these problems represents a rapid radiation that will likely require an increase in both the number of individuals per species and the number of independent nuclear loci. However, vast improvements in understanding carnivore phylogeny have been made in recent years and it is now possible to take this framework and apply it to understanding biogeography and subsequent links to diversification and adaptation. A molecular phylogeny is a piece of the puzzle, along with morphology and ecology, in understanding speciation and evolution in general. Fossil taxa can

also be incorporated to examine speciation patterns through time. For example, further biogeographic work may lead to an understanding of why historically speciose and diverse families such as the walrus and red panda families are now reduced to a single extant species. Conversely, why are close relatives of these monotypic families, the otariids and mustelids, so diverse in the present? What are the environmental, genetic, and ecological conditions that lead to diversifications such as true seals in either polar region or mustelids worldwide? If the conditions that lead to speciation and extinction can be identified, this information can be used to guide conservation efforts toward regions of the world that exhibit the potential for diversification, be it in their environment, species composition, or other yet-unidentified factors.

From a more general perspective, the problems in estimating arctoid phylogeny and the applications of a resolved phylogeny are applicable across all life. Problems of missing data (Chapter 2), phylogenetic artifacts (Chapter 3), and gene incongruence and deep coalescent events (Chapters 4, 5) often only increase as studies move deeper back in time with problems of paralogy or other factors such as horizontal gene transfer and speciation through hybridization more common in bacteria or plants, but still present in many vertebrates. Through this work, I have identified various strengths and weaknesses of several analysis methods and highlighted some of the benefits and problems of combining multiple loci. The conditions under which phylogenetic reconstruction methods fail still requires a great deal of work, as do methods to identify when artifactual topologies are being recovered in empirical studies when the true tree cannot be known. The incorporation of uncertainty in phylogenetic reconstruction and alternate methods of illustrating evolutionary histories beyond bifurcating topologies need to continue to improve, but until we can fully understand how and why evolution and speciation occurs, we remain unable to properly model such processes. Each resolved phylogeny provides a step toward understanding evolution and the potential for improvement into evolutionary models of adaptation, molecular evolution, and speciation. If using speciation patterns to create improved models of molecular evolution can help to understand adaptation and the many other factors involved in the process of speciation, understanding evolution becomes possible.

**References**


Blanquart, S., Lartillot, N., 2006. A bayesian compound stochastic process for modeling nonstationary and nonhomogenous sequence evolution. Mol.Biol.Evol. 23, 2058-2071.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics 2, e68.

Galtier, N., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc.Natl.Acad.Sci. 92, 11317-11321.

Higdon, J.W., Bininda-Emonds, O.R.P., Beck, R.M.D., H., F.S., 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a mutigene dataset. BMC Evol.Biol. 7, 216.

Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, A.W.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimates. Syst.Biol. 53, 638-643.

Koepfli, K.-P., Deere, K.A., Slater, G.J., Begg, C., Begg, K., Grassman, L., Lucherini, M., Veron, G., Wayne, R.K., 2008. Multigene phylogeny of the Mustelidae: Resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. BMC Biology 6.

Koepfli, K.-P., Gompper, M.E., Eizirik, E., Ho, C.-C., Linden, L., Maldonado, J.E., Wayne, R.K., 2007. Phylogeny of the Procyonidae (Mammalia: Carnivora): molecules, morphology and the Great American Interchange. Mol.Phylogenet.Evol. 43, 1076-1095.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concantenated data under coalescence. Syst.Biol. 56, 17-24.

Liu, L., Pearl, D.K., 2007. Species trees from gene trees: Reconstructing posterior distributions of a species phylogeny using estimated gene tree distributions. Syst.Biol. 56, 504-514.

Lockhart, P.J., Novis, P., Milligan, B.G., Riden, J., Rambaut, A., Larkum, T., 2005. Heterotachy and tree building: a case study with plastids and eubacteria. Mol.Biol.Evol. 23, 40-45.

Nakagome, S., Pecon-Slattery, J., Masuda, R., 2008. Unequal rates of Y chromosome gene divergence during speciation of the Family Ursidae. Mol.Biol.Evol. 25, 1333-1343.

Pages, M., Calvignac, S., Klein, C., Paris, M., Hughes, S., Hanni, C., 2008. Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. Mol.Phylogenet.Evol. 47, 73-83.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., Delsuc, F., 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol.Biol. 5.

Ruano-Rubio, V., Fares, M.A., 2007. Artifactual phylogenies caused by correlated distributions of substitution rates among sites and lineages: the good, the bad, and the ugly. Syst.Biol. 56, 68-82.

Wynen, L.P., Goldsworthy, S.D., Insley, S.J., Adams, M., Bickham, J.W., Francis, J., Gallo, J.P., Hoetzel, A.R., Majluf, P., White, R.W.G., Slade, R.W., 2001. Phylogenetic relationships within the eared seals (Otariidae: Carnivora): Implications for the historical biogeography of the family. Mol.Phylogenet.Evol. 21, 270-284.

Yu, L., Li, Y.-W., Ryder, O.A., Zhang, Y.-P., 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol.Biol. 7, 198.

Zhou, Y., Rodrigue, N., Lartillot, N., Philippe, H., 2007. Evaluation of the models handling heterotachy in phylogenetic inference. BMC Evol.Biol. 7, 206.